# Essays in Empirical Asset Pricing

# Essays in Empirical Asset Pricing

Essays over het empirisch prijzen van financiële producten

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Wednesday 22 November 2023 at 10:30 hrs

by

**Amar Soebhag**

born in Rotterdam, The Netherlands

Erasmus University Rotterdam

**Doctoral Committee**

**Promotor:** Prof. Dr. G. Baltussen

**Other members**: Prof. Dr. S. van Bekkum

Dr. E. Eiling

Dr. M. Szymanowska

**Co-promotor:** Prof. Dr. P. Verwijmeren

Tinbergen Institute, Erasmus School of Economics dissertation series number: 830
ISBN 978-90-5892-677-7
© 2023, Amar Soebhag

Design: PanArt, www.panart.nl
Print: OBT bv, www.obt.eu

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the author.

This publication (cover and interior) is printed on FSC® paper PlanoSuperior.

FSC www.fsc.org
MIX
Papier van verantwoorde herkomst
FSC® C022586

**ACKNOWLEDGEMENT**

Doing a PhD is like travelling the world, with each chapter signifying a different destination. During the past four years, I have not been travelling alone, and those years went by in the blink of an eye: Good company on a journey makes the road seem shorter! Hence, I would like to take the opportunity to thank those who have been part of my journey.

First, I would like to thank to my promotor, Prof. Dr. Guido Baltussen. We met 6 years ago, when I followed his seminar "Behavioural Investing". I discovered my passion for data-driven research in finance during his seminar. After the seminar, I wrote my master thesis under his supervision, and decided to continue with a research master at the Tinbergen Institute, and subsequently my PhD under his supervision. I am grateful to Guido for his invaluable advice, and continuous support during my PhD. His immense knowledge and experience encouraged me to lift my own work to the next level. I'm looking forward to our collaborations in the future.

In addition, I would like to show my gratitude to my second promotor, Prof. Dr. Patrick Verwijmeren. Patrick was always willing and enthusiastic to assist in any way he could throughout the PhD.

Furthermore, I want to thank my colleagues at the Business Economics department of the Erasmus School of Economics for providing a pleasant working environment in the past four years. I enjoyed all the seminars, and research discussions. All my colleagues at the department were always very open, helpful, and eager to provide feedback. I also want to thank my colleagues at the Quant Fixed Income department, which I joined in February 2023. I couldn't wish for a more amazing team.

Moreover, I would like to thank my co-authors. First of all, Bart van Vliet for being an amazing co-author, a great investment professional, and a very bright PhD-student. We had countless discussions on research ideas, leading to chapter 3 of my dissertation, and hopefully more research output in the future. I would also like thank Pim van Vliet, Head of conservative investing at Robeco and a co-author (on a low-volatility paper), and our discussions on the low-volatility effect. Last but not least, I'd like to thank Esad Smajlbegovic (ESE), whom I am currently working with on multiple exciting projects.

I am very grateful for the support from my friends, family, and my girlfriend Nadia. My girlfriend Nadia has supported me in each stage of the PhD. I cannot thank her enough for providing the time (many evenings and weekends) and space to work on this dissertation. Lastly, I thank my loving parents for their unconditional love, care, and support. Some of their life lessons are deeply ingrained in me, and have shaped me into who I am today.

Place: Rotterdam

Date: 16/07/2023 **Amar Soebhag**

# Contents

# Chapter 1

# Introduction

Empirical asset pricing, as a prominent sub-field within finance, has received substantial attention in academia and the financial industry over the past decades. Its importance lies in providing insights on the drivers of asset returns, as well as guiding investors, portfolio managers, and policymakers in their decision-making process. By attempting to explain the relationship between risk and return, asset pricing models offer essential frameworks for understanding the pricing of various financial instruments, such as stocks, bonds, and more complex products, like derivatives.

The classical view on asset pricing assumes that investors form rational homogeneous expectations and have full information. Rational expectations theory assumes that individuals form their expectations on future economic variables and asset prices in a rational and unbiased manner. Rationality plays a crucial role in financial theory, since it implies that economic agents do not systematically make forecasting errors. Full information implies that economic agents have access to complete and accurate information about the assets being traded, as well as all relevant economic variables and factors that may affect the assets' prices. These two key assumptions led to the emergence of foundational financial theories in the mid-20th century, such as the Efficient Market Hypothesis by Eugene Fama in 1964.

While rational expectations and full information are elegant and intuitive when modeling various financial phenomena, real-world markets and human behaviour may not always meet these ideal conditions. Behavioural finance acknowledges that market participants may not always act rationally. As such, behavioural finance challenges the traditional view by significantly departing from the traditional asset pricing framework. For example, the literature on behavioural finance documents investors are subject to a wide-range of cognitive biases, heuristics, and emotions that influence their decision-making, leading to systematic misjudgments. Consequently, asset pricing may not always accurately reflect fundamental values, leading to mispricing and market inefficiencies. Furthermore, investors have limited attention and cognitive capacity to process all available information, leading to underreaction to new information, causing assets to become mispriced. Furthermore, during the past three decades, the stream of literature on empirical asset pricing has identified hundreds of market anomalies that can't be explained by traditional asset pricing models. Lastly, non-information based trading is becoming increasingly prevalent in financial markets in recent years. Various types of agents may trade at different times during the day, driven by distinct motivations and information sets, resulting in predictable order flow and systematic price pressures. This challenges the traditional framework which assumes investor homogeneity. The field of asset pricing has changed tremendously compared to the 60s, with traditional paradigm shifting towards the integration of behavioural finance. Furthermore, the growth in computing power and data availability led to an increasing adoption of quantitative and empirical approaches in asset pricing. The dynamic environment, however, also gives rise to new challenges in asset pricing research.

This dissertation focuses on presenting four distinct papers that make valuable contributions to the empirical asset pricing literature. Each paper delves into a different challenge within this field, yet they all share a common thread of studying investor behavior and market inefficiencies. Chapter 2 studies how economic forecasters and markets react to aggregated macro-economic news, which allows to study the rationality, and expectation formation. Chapter 3 addresses a recent challenge in the literature: p-hacking in academic research. In the past three decades hundreds of new factors that arguably explain the cross-section of returns have been discovered. The p-hacking literature recognizes that not all discoveries are valid, but some may be driven by luck or p-hacking. This chapter adds a new dimension to the p-hacking literature: freedom in research choices may allow for p-hacking. Chapter 4 touches upon non-information based trading, where we show that the rebalancing of option market makers is negatively priced in the cross-section of equity returns. Chapter 5 studies intraday return patterns and links the predictability to specific investor types, thereby challenging the investor homogeneity assumed in the traditional framework. We introduce each chapter in more detail below:

In Chapter 2, titled "Caught by Surprise: How Markets Respond to Macroeconomic News," which is a collaborative effort with my PhD advisor, Guido Baltussen, we investigate the impact of macroeconomic surprises on asset prices at the market level. Macroeconomic news plays a pivotal role in shaping the political landscape and influencing financial markets. Remarkably, some savvy investors have demonstrated the ability to capitalize on macroeconomic news and historically outperform the market. The renowned investment figure, David Harding, Founder of Winton Capital Management, eloquently expressed this phenomenon when he said, "We tend to make money out of surprises.. Most surprises unfold gradually." His statement underscores the significance of unexpected events in driving market movements and the opportunities they present for skilled investors to generate returns. Throughout this chapter, we delve into the complexities of how markets respond to these surprises, reveal the underlying mechanisms and implications for asset pricing.

We are raising two questions in this chapter: How can we systematically measure macroeconomic surprises, and how do these surprises unfold in terms of market movements? In this chapter, we study the macroeconomic news flow across major regions. Macroeconomic news are released on a, practically, daily basis. Many professional forecasters offer frequent projections on the future path of well-followed macroeconomic news series. However, such projections are imperfect from the actual realizations, leading to surprises in macroeconomic releases. We introduce a PCA-based methodology to model the surprises in the daily flow of macroeconomic news, consequently creating a daily macro-economic surprise index. Our proposed methodology allows to deal with a large number of surprises series (we use over 200 series ourselves) that have different frequencies as well as different release times. We find that surprises show momentum-like behaviour, driven by anchoring biases. As such forecasters seem to underreact to macroeconomic news, and slowly update their beliefs. We find a similar underreaction when we study the predictive power of the surprise index on future asset returns. Our surprise index positively and significantly predicts future returns across different markets and asset classes.

In Chapter 3, our research is motivated by the ongoing discussions surrounding p-hacking and the replication crisis in finance. Over the past decades, the number of factors purported to predict returns in the cross-section has grown significantly. However, critics argue that some of these factors may be spurious, arising from luck or even intentional manipulation of methodologies and data to yield high t-statistics. The pressure to publish statistically significant results may also incentivize researchers to engage in p-hacking. To address these concerns, recent literature proposes innovative econometric methods to distinguish between "true" factors and those that are merely "lucky or false". Additionally, there is a call for adopting code - and data-sharing practices to enhance transparency and facilitate the replication of findings. Furthermore, researchers are encouraged to employ more conservative methods when identifying new factors, ensuring that only robust and meaningful factors make their way into asset pricing models.

This discussion inspired Bart van Vliet, Patrick Verwijmeren, and me to work on a study that by now is evolved into chapter 3 of this manuscript. We add an additional argument in the discussion: researchers face a number of methodological decisions when engaging in research. While it allows for research customization, it also leaves room for p-hacking. Specifically, in this study, we focus on the portfolio sorting methodology, which is the most used method within empirical asset pricing. We reviewed over 300 articles published in this field that use portfolio sorting, and find that more than 10 different decisions are involved in portfolio sorting. Interestingly, we document a large variation in research choices made in these articles. Subsequently, we argue that variation in choices may add to variation in outcome. We select multiple well-known factor models, and construct them in every possible combination, given a set of eleven research choices, leading to 2048 construction combinations. Our results suggest variation in outcomes due to research choices is as important as statistical significance.

In chapter 4, a solo-authored study, we focus on option markets. Option volumes have been increasing since the introduction of exchange-based option trading in 1973. In recent years, the increase has accelerated, partially driven by increased participation of retail investors. Option trading volumes have surged to such high levels that it even surpassed stock volumes on some days. A period worth mentioning was the beginning of 2021, when bubbles were formed on certain micro-cap stocks, such as GameStop (GME) and Bed, Bath, and Beyond (BBB). These companies experienced an astronomical surge in stock prices, and this phenomenon were partly attributed to a substantial surge in options trading associated with these stocks.

This led us to ask a key question: does option trading impact the return dynamics of its underlying asset? According to the Black-Scholes model, the simple answer is "no". The model postulates no feedback effect from option prices to stock prices. However, we argue that a feedback effect may exist in financial markets. The idea relies on the notion that option market makers, by mandate, remain delta-neutral, i.e. they hedge away exposure towards the underlying. Option deltas measure how much an option market maker needs to hedge, but these are a function of the stock price itself. Option gammas measure how aggressively option deltas change when stock prices fluctuate. In other words, option gammas can be used as a proxy to measure how aggressively an option market maker needs to hedge. When the amount that needs to be hedged away is large, this can create additional price pressure beyond fundamental news. In this paper, we construct a stock-level measure of option market making hedging, and show that it is negatively priced in the cross-section of U.S. equity returns. We show that this effect is robust, and is driven by non-informational trading.

In recent years, high-frequency research is becoming more prevalent. With the increase in computational power, and data storage, analyzing high-frequency data becomes easier. In chapter 5, we (Guido Baltussen and I) study high-frequency return predictability of U.S. stock returns over a 27-year sample. Different market participants systematically trade various kinds of assets for different reasons and at different times during the trading session. Studying high-frequency return patterns provides insights on the behaviour of such systematic order flow. Previous research has shown that intraday predictability exists at the market-level, whereby the return before the last half-hour predicts the last half-hour return positively. This effect is driven by hedging demand by option market makers, an effect that we also studied in the previous chapter. We extend this line of research by also studying intraday returns at the stock-level. In contrast to market returns, we find that stock-level returns show the exact opposite pattern: intraday reversal. Stocks that tend to do well under-perform during the last half-hour relative to stocks that did not do well before.

We find that this effect is not only driven by hedging demand by option market makers, as is shown in previous studies. In fact, we argue that there are two other forces playing a major role at the end of the day: retail traders, and short sellers. Retail traders typically are contrarian traders relative to institutional investors. We find that retail traders engage in dip-buying behaviour around the end of the day, which pushes the stock price back up again. Short sellers tend to close positions around the end of the day, which also yields additional price pressure at the end of the day. Furthermore, we show that intraday reversal is a distinct pattern compared to other well-known systematic intraday return patterns.

The insights gained from these chapters also have wide-ranging implications for academics, investment professionals, and policymakers. Chapter 2 shows that forecasters and investors underreact to macroeconomic news, inconsistent with the traditional framework. Furthermore, we offer a method to conduct economic nowcasts. Real-time information, via nowcasting, can have substantial impact on financial markets, as we show. Investors can use these nowcasting techniques to make informed decisions on asset allocation, risk management, and market positioning. For policymakers, economic nowcasting helps to monitor the state of the economy and may aid in identifying potential vulnerabilities in the economy that require policy intervention. Chapter 3 highlights that freedom of research choice can be used for p-hacking purposes. Both for academic researchers and investment professionals, we provide several suggestions and recommendations to limit p-hacking. Chapter 4 highlights the existence of feedback effects from options to stocks, inconsistent with the Black-Scholes option pricing model. Option market makers rebalance their inventory in order to remain delta-neutral, by mandate. We measure how aggressively an option market maker must rebalance via its net gamma exposure, and show that this negatively priced in the cross-section of stock returns. For both academia and investment professionals, we highlight the importance of information contained in option prices and characteristics. For policymakers, monitoring net gamma exposure can be of added value in a risk management framework, given that we also show that net gamma exposure significantly predicts future realized stock volatility. Chapter 5 provides insights on a new intraday pattern: intraday reversal, whereby the last-half hour return reverts relative to the preceding return during the trading day. We show that the effect is not only driven by price pressure induced by hedging demand, but also by retail traders and short-sellers during the last half-hour. From an academic perspective, our results challenge investor homogeneity as assumed in the traditional asset pricing framework. For investors, we highlight that the last half-hour contains unique dynamics compared to other return intervals. This insight can be, for example, used to trade more efficient and reduce transaction costs.

Chapter 6 provides a summary of this manuscript. This summary is available in English and in Dutch.

## Declaration of contribution

**Chapter 2**: This chapter is based on joint work with Guido Baltussen. The study has benefited from Patrick Verwijmeren, Zhi Da, Andre Lucas, Bas Peeters, Bart van Vliet, Vadym Volosovych, Charlie Clarke, seminar participants at the European Finance Association, Erasmus School of Economics, Robeco Institutional Asset Management, Tinbergen Institute, Erasmus Research Institute of Management, and Lancaster University. The initial idea and concept were proposed by Guido Baltussen. Literature review, data collection, and empirical work has been conducted by me. The writing is joint with Guido Baltussen.

**Chapter 3**: This chapter is based on joint work with Bart van Vliet, and Patrick Verwijmeren. The study has benefited from helpful comments and feedback from Guido Baltussen, Pedro Barroso, Victor DeMiguel, David Feldman, Paul Fontanier, Peter Koudijs, Frederik Muskens, Dimitris Papadimitriou, Esad Smaljbegovic, Sjoerd van Bekkum, Remco Zwinkels, seminar participants at Robeco Institutional Asset Management, Tinbergen Institute, Erasmus Research Institute of Management, and conference participants at the 2022 Finance Symposium, 2022 Frontiers of Factor Investing conference, and 2023 Portuguese Finance Conference for fruitful discussions and feedback. The initial idea and concept were developed by me. Data collection, and empirical work has been conducted by Bart van Vliet, and me. The literature review and writing is joint with Bart van Vliet, Patrick Verwijmeren and me.

**Chapter 4**: This chapter is solo-authored work by me. The work has benefited from comments and feedback from Guido Baltussen, Patrick Verwijmeren, Bart van Vliet, seminar participants at Robeco Institutional Asset Management, Tinbergen Institute and Erasmus Research Institute of Management, and conference participants at the 2023 Financial Management Association, European Meeting in Aalborg, Denmark. Each stage of the research process has solely conducting by me: hypothesis development and research question, writing, literature review, data collection, and empirical work.

**Chapter 5**: This chapter is joint work with Guido Baltussen. The work has benefited from comments and feedback from Patrick Verwijmeren, Bart van Vliet, seminar participants at Robeco Institutional Asset Management, PGGM Quant Equity Research, Tinbergen Institute and Erasmus Research Institute of Management. Literature review, data, and empirical work is done by me. Hypothesis development and writing is joint with Guido Baltussen.

# Chapter 2

# Caught by Surprise: How Markets Respond to Macroeconomic News[1]

## 2.1 Introduction

Macroeconomic news is released in markets practically every day, with a big industry analyzing current and forecasting upcoming news. The unexpected component of this information, or surprise, typically causes investors to update their expectations and trigger investment decisions. Several successful investors even claim to profit from the gradual unfolding of macroeconomic surprises.[2] Yet, relatively little is known about the behavior of macroeconomic surprises and their incorporation in asset prices. In this paper we comprehensively study macroeconomic surprises across three key dimensions. First, we develop a novel aggregate nowcast metric of surprises across hundreds of macroeconomic series. Second, we investigate biases in macroeconomic forecasts and the resulting behavior of macroeconomic surprises globally. Third, we examine their incorporation into asset prices across major asset classes.

Three major challenges arise in studying the impact of macroeconomic surprises. First, hundreds of individual macroeconomic series exist. Almost daily, macroeconomic numbers are released that are studied and appreciated by investors. Previous studies have typically focused on a few selected macroeconomic series over their release frequency, such as the PPI, CPI, non-farm payrolls, or the unemployment rate (see amongst others Boyd, Hu, and Jagannathan (2005), Smirlock (1986), Urich and Wachtel (1984), and Beber and Brandt (2009)). However, many more are appreciated by investors as important economic news, with in the U.S. alone over 50

[1]This chapter is based on Baltussen and Soebhag (2022).

[2]For example: "We tend to make money out of surprises... Most surprises unfold gradually" - David Harding, CEO Winton Capital Management, taken from Pedersen (2019)

well-followed releases every month. Second, macroeconomic series are released relatively infrequent, with the most common release frequency being monthly. As a consequence, individual macroeconomic releases are observed at a relatively low frequency, unlike the near-continuous flow of macroeconomic information available to investors. Third, macro news needs to be measured as was historically available to investors. Many macroeconomic series are restated after their initial release and are commonly released with a delay. Capturing information in real-time is crucial, as Ghysels, Horan, and Moench (2014) find that the use of non-restated data and properly accounting for publication delay undoes most of the previously found bond return predictability using macroeconomic information by Ludvigson and Ng (2009). Importantly, macroeconomic news also continues after the initial release of a series in its revisions.

To tackle these challenges, as our first major contribution, we develop a novel real-time nowcast metric of macroeconomic surprises across hundreds of macroeconomic releases. To this end, we extend the method introduced by Beber, Brandt, and Luisi (2015) to macroeconomic surprises. Their method allows to distill the news flow of dozens of macroeconomic releases observed at different times and frequencies into a daily measure of aggregate macroeconomic conditions by extracting the latent factors across a large set of macroeconomic series within pre-specified economic categories based on the cross-sectional commonality.[3] Following Beber et al. (2015), we classify series into two main categories of macroeconomic fundamentals, namely economic growth and inflation, generally seen as the main macroeconomic drivers of financial markets. We combine the weights resulting from this method with the latest surprise of individual macroeconomic series to compute aggregate macroeconomic surprises. The result is real-time nowcast measures of macroeconomic surprises that summarizes the wealth of macroeconomic releases into a single reading on growth or inflation, which can be utilized to understand the aggregate behavior of macroeconomic surprises and its incorporation in asset prices.

The main benefits of the introduced method are that it (i) allows for aggregating a large number of macroeconomic releases that are relevant for tracking economic conditions into a single number, (ii) focuses on well-interpretable categories of economic information (growth and inflation), instead of more statistical factors, (iii) relies on real-time information (not utilizing future information), thereby not creating a confounding impact on for example future returns, and (iv) offers a measure of macroeconomic surprises available at the daily frequency.[4] Moreover, our method

[3]Beber et al. (2015) shows that their resulting real-time measure is highly correlated with other approaches, like GDP, CFNAI and the Aruoba, Diebold, and Scotti (2009) business condition index (ADS), but appears more timely and informative about future macroeconomic fundamentals.

[4]Other approaches that summarize the cross-section of macroeconomic series into a few common dimensions are, amongst others, the dynamic factor approach of Ludvigson and Ng (2009), the CFNAI index based on Stock and Watson (1989), and the approaches of Aruoba et al. (2009) or Gilbert et al. (2017) (GSSV). The dynamic factor approach extracts the main latent factors, complicating their exact interpretation as either growth or inflation factors, and is not available at the daily frequency. CFNAI is only available at monthly frequency. Aruoba et al. (2009) only considers a few indicators, and include market-based data, which may already reflect the market's interpretation of

is simple to implement, aggregates a large number of releases available at different times and frequencies, is well-interpretable, and can easily be extended to other markets, categories, or frequencies.

As an additional differentiating feature of our approach, we compose a surprise from two parts: an 'instantaneous announcement surprise' and a 'revision surprise'. The first part captures the surprise to the current number, and the second part the revision to the past number. For example, the U.S. non-farm payrolls (NFP) number over December 2021, released January 07, 2022, was 199k, the forecast at that time was 450k, and on February 04, 2022, the December number was revised to 510k. Consequently, the instantaneous announcement surprise equals -251k (199k - 450k), while the revision surprise equals +311k (510k - 199k). Both parts reflect the arrival of new information, as we subsequently confirm in our empirical analysis (see also Krueger and Fortson (2003), Faust et al. (2007), Croushore and Stark (2003), Croushore (2011), and Gilbert (2011) for evidence that revisions can be substantial and important).

Our second major contribution is to comprehensively study the behavior of macroeconomic surprises, documenting a strong stylized fact. Our sample runs from March 1997 (the start date of our macroeconomic forecast data) to December 2019 across four major economic regions (U.S., U.K., the Eurozone, and Japan) and spans most macroeconomic series available over this sample. Our findings show that economic surprises do not follow a random walk but display sizable short-term autocorrelation over the next months. We term this empirical pattern 'economic surprise momentum'. This momentum is present across all regions, especially present in the growth category, and correlated over regions. We conjecture that economic surprise momentum occurs systematically due to (i) underreaction in economists' consensus forecasts to the series' time-series behavior, akin to the behavior found in analyst forecasts about short-term firm fundamentals (Bouchaud et al. (2019)), and, novel to the literature, (ii) underreaction in consensus forecasts to information in surprises of other series. Our test confirm this conjecture. First, changes in forecasts are too rigid compared to actual changes and anchored towards previous releases. Second, when we decompose economic surprise momentum into autocorrelation in individual series and cross-autocorrelation between series, in spirit of autocorrelation decomposition in returns proposed by Hou (2007) and Baltussen, van Bekkum, and Da (2019), we find that economic surprise momentum in growth variables originates for 50% to 60% from autocorrelation in individual series, and 40% to 50% from cross-autocorrelation across macroeconomic series. In other words, a past surprise in, for example, non-farm payrolls predicts a future surprise in non-farm payrolls, but also in other macroeconomic series like the unemployment rate and jobless claims.

economic news. The GSSV approach assigns weights to each macroeconomic release depending on the contribution to a business condition index capturing the future state of the economy. Related, Scotti (2016) proposes a dynamic factor method to summarize macroeconomic information of a handful key indicators of real activity based on their contribution to a coincident index of economic activity (the ADS index). By contrast, we use much more (relevant) macroeconomic information, apply a simple and straightforward method, and include both announcement and revision surprises.

Our results on consensus forecasts align with results in previous studies on macroeconomic forecasts, generally documenting deviation from the full information rational expectations hypothesis in the direction of underreaction in consensus forecasts. Most notably, Coibion and Gorodnichenko (2012, 2015) and Bordalo, Gennaioli, Ma, and Shleifer (2020) (henceforth BGMS) find evidence of underreaction in the consensus forecasts of several key macroeconomic variables.[5] These studies have a different focus as ours, as they examine predictability of forecast errors from revisions to forecasts of the same release over time. Instead, we examine the relationship between surprises and consensus forecasts made just before before macroeconomic announcements across hundreds of macroeconomic forecasts across the globe. Interestingly, underreaction in consensus forecasts is present in forecast revisions and in forecasts just before macroeconomic announcements. Further, we show that consensus level underreaction is driven by two forces; underreaction to a series' time series behavior and underreaction to the information contained in recent surprises on other macroeconomic series. The latter is novel to the literature, and aligns both with the models of Coibion and Gorodnichenko (2012, 2015) and BGMS assuming forecasters have more inattention to other series. As such, we extend the findings of existing studies to information spillover across macroeconomic releases and to a large panel of forecasts across the globe taken just before macroeconomic releases.

As our third major contribution, we examine the incorporation of economic surprises into asset prices. Macroeconomic forecasts are heavily followed in markets, and hence an important question is whether autocorrelation in macroeconomic surprises is also reflected in asset prices? In an efficient market, macroeconomic surprises should not systematically predict asset returns with the exception of the incorporation of new information directly around the release (Fama et al., 1969). As such, one could conjecture that rationality of macroeconomic forecasts can be tested by the efficient incorporation of its information into asset prices. Arguably, if economic surprise momentum is a reflection of rational informational frictions, investors should not be able to profit from these frictions. By contrast, if expectational errors in macroeconomic forecasts impact investment decisions, surprises would predict asset returns. As such, examining return predictability provides a natural test on rational versus irrational explanations of systematic patterns in economic surprises.

We examine the predictability of returns across all major asset classes (equities, credits, commodities and government bonds) and across the major regions (U.S., U.K., the Eurozone and Japan) over our full sample period. We find that asset returns are strongly predictable by macroeconomic surprises using both traditional predictive regressions and trading-based investment strategies. Our macroeconomic

[5] At the level of the individual forecaster (instead of the consensus level), BGMS and Broer and Kohlhas (2018) document overreaction in most of the macroeconomic forecasts. Interestingly, BGMS attribute these seemingly contradictory findings to overreaction to news at the individual forecaster level combined with information frictions, and show that their interaction determines consensus level over- or underreaction. As we lack data on individual forecasts we have to refrain from examining economic surprise momentum at the forecaster level.

growth surprise nowcast positively predicts returns on risky assets (i.e., equities, credits and commodities), an effect that is economically sizable and statistically highly significant in regressions. Further, simple investment strategies based on growth surprises yield sizable Sharpe ratios and CAPM alphas. Returns on government bonds are predicted negatively, although this effect is less consistent. These findings are robust to control variables previously found to predict returns, and show up across various nowcasting techniques. Studying the horizon of predictability, we find that asset returns tend to be predictable during the next day till three months out. Further, we show that the global macroeconomic surprises are generally more important than regional metrics, and that both the instantaneous announcement surprise and the revision surprise carry predictive information. Overall, we can conclude that economic surprise momentum is also reflected in asset prices, generating economically sizable opportunities.

Finally, we conduct additional tests to show that the return predictability is driven by autocorrelation in macroeconomic surprises. Predictability is generally present when current and future surprises carry the same sign, but tends to be absent otherwise. Further, we examine if the documented return predictability reflects increased risk or risk premia after autocorrelated macroeconomic surprises, but fail to find supportive evidence.[6] Growth surprises negatively predict volatility across asset classes (equity, bonds, credit, and commodities), and yield significantly *negative* expected equity and credit returns after the most negative surprises. A rational time-varying risk premia explanation is hard to align with lower future risk or negative expected returns on risky assets (Baker and Wurgler (2000)). That said, we like to stress that some caution is warranted on these results as risk exposures and risk premia are not directly observable. These findings suggest that the predictability reflects a market inefficiency due to underreaction in investors' expectations, and hence that underreaction is reflected in both published macroeconomic forecasts and asset prices set by investors. Overall, we believe our results show that macroeconomic expectations of forecasters and investors are predictably biased generating economic surprise momentum in macroeconomic news and asset prices.

Our paper contributes to several strands of literature, most notably the voluminous literature on macroeconomic information and asset prices. To date, studies on macroeconomic news have been typically isolated to particular markets or asset classes, a handful of releases, or plagued by substantial measurement issues of macroeconomic fundamentals. For example, announcements on unemployment, balance of trade, housing variables, inflation and money growth have been documented to impact U.S. stock and bond returns, and volatility (McQueen and Roley (1993), Flannery and Protopapadakis (2002), Boyd et al. (2005), Campbell and Diebold (2009), Beber and Brandt (2009), Paye (2012)). Flannery and Protopapadakis (2002) are one of the first to study a broader range of macroeconomic releases for U.S. eq-

[6]Risk premia vary over time and a variable predicting returns on risky assets might indicate predictable variation of risk or its premia across economic states of the world, or alternatively could correlate with changing investment opportunities sets (Merton (1973)).

uities across nominal and real variables (17 in total), a small fraction of all series available. In addition, several papers study the market impact of surprises the minutes to days around the release (Balduzzi, Elton, and Green (2001), Andersen et al. (2007), Faust et al. (2007), Gilbert et al. (2017)), or focus on an announcement relative to an econometric model of expectations (Boyd et al. (2005)). Again, most of these studies focus on one to several announcement series and typically focus on U.S. stock or bond returns. By contrast, we study information in over hundred macroeconomic surprises across all major regions and asset classes, examine asset price responses up to several months ahead, and examine the relevance of both global and regional macroeconomic news.

Our paper also relates to the literature on macroeconomic nowcasting, an increasingly popular technique to forecast the current state of the economy in real-time. In this regard, our paper is most closely related to Beber et al. (2015), but differs in several important aspects. Most importantly, we extend their method to macroeconomic surprises, and analyse the behavior and market impact of these surprises, whereas Beber et al. (2015) focus on the level of macroeconomic conditions. Further, our analysis is more comprehensive in its coverage by spanning all major regions and asset classes, and we include several additional macroeconomic releases. But, most importantly, we discover a strong and robust stylized fact: economic surprise momentum.[7]

Finally, our results add to the growing literature on expectational errors. La Porta (1996) and Bordalo et al. (2019) document misreaction in the earnings growth expectations of analysts, an effect priced into the cross-section of stock returns. Landier, Ma, and Thesmar (2019) find evidence of overreaction in forecasts about macroeconomic processes in controlled experiments. Beshears et al. (2013) show that experimental subjects fail to correctly assess the dynamics of processes that exhibit short-term momentum and long-run mean reversion, processes that characterize the behavior of macroeconomic variables like GDP, unemployment and corporate earnings.

The remainder of the paper is structured as follows. In section 2.2, we introduce our novel nowcast method for macroeconomic surprises. Section 2.3 examines the dynamics of individual and aggregate macroeconomic surprises. Section 2.4 investigates return predictability by macroeconomic surprises, including the role of revision surprises, alternative aggregation techniques, and other robustness tests. Section 2.5 explores the channels behind the predictability. Finally, section 2.6 concludes.

[7]Related, Chousakos and Giamouridis (2020) apply the approach of Beber et al. (2015) to predict asset returns in the cross-section, finding a sizable and significant economic growth premium.

## 2.2 Nowcasting macroeconomic surprises

In this section we describe our novel real-time nowcast metric of macroeconomic surprises. Every month, hundreds of macroeconomic numbers are released across the globe, typically at different times and frequencies. To study the impact of macroeconomic surprises across markets, we believe it is important to comprehensively capture the macroeconomic information flow across all major releases (released at different times and frequencies), use real-time information as actually released on the announcement days (as opposed to restated macroeconomic data), and combine into a daily measure of the current reading of macroeconomic surprises. To this end, we develop a novel real-time nowcast of macroeconomic surprises across (almost) all macroeconomic releases, classified into the growth and inflation state of the economy for the four main economies: U.S. (US), the Eurozone (EU), U.K. (UK), and Japan (JP). This surprise nowcast summarizes the information of macroeconomic releases into a single reading on growth or inflation surprises, which can be utilized to understand the aggregate behavior of macroeconomic surprises and its incorporation in asset prices. The methodology takes the nowcasting approach recently introduced by Beber et al. (2015) as starting point, and extends it to macroeconomic surprises. Below we describe the methodology and data in detail.

### 2.2.1 Real-time macroeconomic information

To have maximum power in our analyses we build a sample that runs from March 1997 (the start date of our macroeconomic forecast data) to December 2019 across the four major economic regions (U.S., U.K., the Eurozone, and Japan) and spans all macroeconomic series with sufficient data coverage over this sample. We collect real-time macroeconomic announcement records from the Bloomberg Economic Calendar (BEC). Since 1997, BEC provides time-stamped, non-revised and non-restated data on hundreds of macroeconomic series. Each announcement record consists of the release date and time, actual released announcement value, consensus forecast (defined as the median expected release value for the upcoming news announcement across individual economist forecasts), standard deviation of the individual forecasts, the number of individual forecasts, and the revision to the previous announcement value.[8] Bloomberg consensus estimates are widely followed by investors, commonly considered as being a major source of economists expectations, as for example witnessed by their frequent appearances in headline news and by actual numbers deviating from these numbers triggering a market reaction (Vrugt, 2009).

We select all headline macroeconomic releases that have data available, are not financial market data (most notably interest rate decisions), and are well tracked, assessed via a Bloomberg relevance score of 25 or above, and interviews with professional investors or economists. This index tells us how closely an announcement

[8]Bloomberg forecasts are based on surveys amongst professional economists during the weeks up to three days prior to the announcement date.

is followed, predominantly by professional investors, and hence signals the importance of the macroeconomic announcement to investors.[9] In addition, we require macroeconomic series to have at least three years of announcement data to allow for an efficient incorporation into our aggregation methodology. We ignore financial market measures related to interest rates or central bank policy to focus on macroeconomic news announcements not driven by market or interest rate movements, as such information may already reflect market's opinion on the state of the economy. In contrast, several other macroeconomic studies use information like interest rates, credits spread, stock prices or the VIX index (Ludvigson and Ng (2009), Aruoba et al. (2009)).

This selection results in 62 distinct macroeconomic series for the U.S. (compared to 43 for Beber et al. (2015)), 72 for the EU, 27 for the U.K. and 30 for Japan.[10] Globally, we use almost 200 different macroeconomic series. This amounts to a dataset containing approximately 13,500 unique surprises for US, 4,800 for UK, 4,700 for Japan, and 11,000 for Europe. Hence, in total our sample includes around 34,000 unique macroeconomic surprises.[11] Our set of selected macroeconomic variables extends the list of Beber et al. (2015) by adding a number of housing related releases (such as home sales). These variables are widely tracked by investors and economists, and provide additional information about the state of the economy. In addition, we add several other macroeconomic variables that score high on the Bloomberg relevance index. In the appendix, we outline each macroeconomic news series that is included in our sample.

Bloomberg also contains data prior to 1997. However this information is stored in historical fields dated according to their reference period. In addition, this information may have been restated over time. As such we start our sample in 1997, but collect the historical data before 1997 to construct an estimate for a initial correlation matrix, as will be explained in section 2.2.2.

### 2.2.2 Data transformations

We apply transformations to several time series with announcement values in order to make them stationary. To check for stationary, we conduct an (augmented)

[9]Bloomberg's relevance score represents the number of alerts set on Bloomberg terminals for an economic event relative to all alerts set for a particular country or region.

[10]We focus on the headline series of an economic report, ignoring sub-series released in the same report. Further, some specific remarks based on specialities in the series selection per region. US: we exclude the Conference Board US Leading Index as this is a composite index which is already covered by other variables which we already include. EU: we only choose variables that are from Germany, France, Italy, Spain and the Eurozone aggregate, as these countries cover the major part of the Eurozone economy.

[11]Several series have multiple releases of the same announcement, for example GDP. We choose to only use the final releases as of their release date, ignoring first or preliminary releases. Our method can be extended to also take these into account, which we leave to future work as we do not have their historical data available.

Dickey-Fuller test for each macroeconomic variable. Combined with the description and definition of a variable, we determine whether a series is stationary from an economic perspective, akin to Beber et al. (2015). A growth rate, for example, would be considered as a stationary time-series. Time-series that are non-stationary are first-differenced before the start of the analysis. Conclusions from the two methods differ in some cases, because some data series are too short with less than 5 years of announcements, for example. In these cases, we rely mostly on the economic definition and description to determine whether the time-series are stationary or not. The appendix lists which announcement series are adjusted to deal with non-stationarity. Furthermore, a small portion of macroeconomic series (mainly U.K. and Eurozone inflation series) are not seasonally adjusted, which we adjust by taking seasoned differences.[12] We have verified that the inflation factors created from seasonally adjusted series are highly correlated with the initial set of data.[13]

Next, we convert all data items from announcement time to calendar time, thereby creating a sparse matrix. Following Beber et al. (2015), we forward-fill the missing value by the last observed value of each macroeconomic series from release to release.[14] In this case, we can think of the time-series in calendar time as a step function that changes when a new announcement is released for that variable. Other statistical models exist to impute missing values, however these models are typically less parsimonious, more complex and less suited to study surprises. Hence, we favor the simplicity of forward-filling.

### 2.2.3 Categorising macroeconomic news

We follow Beber et al. (2015) and impose an economically motivated structure on the macroeconomic news flow. Specifically, we classify our set of macroeconomic announcements into two main categories of macroeconomic fundamentals, namely economic "growth" and "inflation", as these are generally seen as the main macroeconomic drivers of financial markets. We construct daily latent factors that aggregate the macroeconomic series for both categories in our main analyses. This approach has two advantages over other approaches. First, the factors are easier to interpret from an economic point of view, unlike more statistical approaches (e.g. Ludvigson and Ng (2009)), as we essentially follow a 'supervised' approach. Second, it combines all relevant macroeconomic announcements that are studied by investors within each category into one latent factor, in contrast to studies focusing on a single or few announcement time series.

We subdivide the growth category further into three subcategories: output, employment and sentiment. The output category contains information on the supply

[12]We apply a x13-ARIMA model to calculate seasoned differences.

[13]Moreover, results reported in the next sections are qualitatively and quantitatively similar when not applying seasonal differences.

[14]For monthly (quarterly) releases, we forward-fill at most 45 (90) business days.

and demand side of the economy, with series like GDP, industrial production, retail sales, and good orders. The employment category contains information about employment, with series like jobless claims, the unemployment rate, and non-farm payrolls. Sentiment contains subjective views on the current or future state of the economy obtained from surveys, with series like consumer confidence indices, business confidence indices, and purchasing manager indices (PMI's). The inflation category contains series related to nominal-inflation related announcements, with series like consumer prices indices (CPI's) and producer price indices (PPI's). As mentioned before, we ignore financial market measures related to interest rates or central bank policy to focus on macroeconomic news announcements not driven by interest rate movements.

### 2.2.4 Aggregating the macroeconomic news flow

In order to construct real-time high-frequency latent factors, we follow Beber et al. (2015) and use recursive principal component analysis (PCA). We apply this approach on the announcement values and by category and region in order to capture the key common component in its macroeconomic series. Before the start of the PCA, each announcement value time-series is recursively standardized such that it has zero mean and unit variance for every day $t$. To initialize the recursive PCA, we use the historical data between January 1980 and March 1997 and calculate the initial correlation matrix per category. Using this initial correlation matrix, we calculate an initial weight for each macroeconomic series via computing the first principal component. Note that the historical data does not contain the announcement dates, but only the reference dates. We follow Beber et al. (2015) and correct these historical reference date by estimating the median lag length between the reference date and announcement date in the data after 1996. Subsequently, on each day $t$, within category $i$, we first compute the correlation matrix $\Omega_{t,i}$ using information up to day $t$ for all currently active macroeconomic series. We use the correlation matrix, rather than the covariance matrix, to correct for different scaling used for the macroeconomic series. After calculating the correlation matrix at time $t$, we extract the daily factor loadings obtained from the first principal component of this correlation matrix. These loadings are assigned, as a weight, to its corresponding macroeconomic news series within category $i$.

Several technical issues are taken into account when estimating the correlation matrix. First, we have an unbalanced panel data-set, as some macroeconomic series do not span the whole time period. Some series start after the start date of the estimation window or stop to exist before the end of the sample. We take this into account in our methodology by recursively conducting PCA analysis on the cross-section of currently active time-series. Further, some series have records of real-time announcements from an arbitrary date after Jan 1990. We include them into the analysis sequentially, conditional on 3-years of data being available to calculate an initial correlation matrix. To deal with unequal sample lengths, we follow Beber et al. (2015) and use the methodology as proposed by Stambaugh (1997). This methodology results in a correlation matrix estimate that is constructed using adjusted first and second moments. The idea behind the methodology is to use the observed data of the longer time-series with a projection of the shorter series on the longer series, when both are observed, to adjust the moments of the shorter time-series.

Second, most macroeconomic announcement series exhibit serial correlation in calendar time due to the natural autocorrelation of the raw data in announcement time (before forward-filling), misalignment of the news in calendar time, and forward-filling of missing values. Especially forward-filling creates local constancy, leading to high persistence within a time-series. To account for the local constancy, we perform sub-sampling over 21 business days, as described in Beber et al. (2015). On day $t$ we draw 21 sub-samples backwards from the forward-filled news announcement series. These sub-samples start from $t-1$, $t-2$, $t-3$, $\cdots$, $t-21$, respectively (in other words, the first sub-sample has observations on day $t-1$, $t-22$, $t-43$, $\cdots$, etc.). Subsequently, we use the Newey-West method to calculate a heteroskedasticity and autocorrelation consistent correlation matrix using four lags. For each sub-sample, we calculate this correlation matrix, and then take the average over all the estimates of the 21 correlation matrices.[15]

### 2.2.5 Macroeconomic nowcast metrics

We use the factor loadings obtained from the recursive PCA to construct our real-time metric of macroeconomic surprises. More specifically, per category within a region, we use the following formula to build the surprise factors:

$$S^r_{c,t} = \sum_{i \in \omega_t} \lambda^r_{i,c,t} S^r_{i,c,t} \tag{2.1}$$

Superscript $r$ denotes the regions. Subscript $c$ denotes the category. $\omega_t$ denotes the set of currently active macroeconomic news series at day $t$. We normalize the first principal component weights to account for the proportion of missing data. The economic surprise consists ($S$) of two parts. The first part is the instantaneous 'announcement surprise', which is the difference between the latest real-time announcement value and the economists' survey consensus. The second part is defined as the 'revision surprise', defined as the difference between the latest revision value and last period's announcement. This latter term is, to the best of our knowledge, not employed in earlier literature, but does reflect the arrival of new information to investors. Krueger and Fortson (2003), Faust et al. (2007), Croushore and Stark (2003), Croushore (2011), and Gilbert (2011) show that revisions can be substantial and matter to investors. Since we have four regions and four categories, we obtain sixteen macroeconomic surprise factors. We construct the growth metric per region by taking an equal-weighted average of the employment, output and sentiment level

[15] See also Ait-Sahalia, Mykland, and Zhang (2005) for more specifics about this sub-sampling method.

factors.

Overall, the result is an aggregate nowcast metric of macroeconomic surprises that (i) aggregates a large number of macroeconomic releases that are relevant for tracking economic conditions into a single number, (ii) focuses on well-interpretable categories of economic information (growth and inflation), (iii) relies on real-time information, and (iv) offers a measure of macroeconomic surprises available at the daily frequency. In general, the method is simple to implement, can easily handle a large number of releases available at different times and frequencies, is well-interpretable, is available in real-time, and can be easily extended to other markets, categories and frequencies.

In addition, we build the two nowcast metrics developed by Beber et al. (2015): the nowcast of the actual state of the economy ("Actuals"), and the nowcast of the (ex ante) uncertainty, or disagreement, about the actual state of the economy, where we use the announcement value, respectively the standard deviation across individual forecasts, instead of the surprise value:

$$A^r_{c,t} = \sum_{i \in \omega_t} \lambda^r_{i,c,t} A^r_{i,c,t} \tag{2.2}$$

$$FD^r_{c,t} = \sum_{i \in \omega_t} \lambda^r_{i,c,t} FD^r_{i,c,t} \tag{2.3}$$

In our analysis, we are interested in the surprise factors for both the growth and inflation category across the four regions. The level and uncertainty factor are used to control for the actual state of the economy and disagreement among economists.

### 2.2.6 Financial market data and control variables

In order to study whether macroeconomic surprises affect financial markets, we obtain data for several asset classes from Datastream and Bloomberg. For the stock market, we obtain historical prices of the front-month futures contract rolled the day before expiry of the S&P 500, FTSE 100, Nikkei 225 and Eurostoxx 50. For bond markets, we obtain the prices of the front-month futures contract rolled the day before first notice on government bonds with a maturity of 10 years for the U.S., U.K., Japan, and Europe. We scale the bond futures returns to a unit duration position. For credit markets, we obtain returns on 5-year Credit Default Swap (CDS) contracts on investment grade (USIG) and high yield (USHY) corporate bond indices for the U.S., as well as corporate bond investment grade (EUIG) and high yield (EUHY) indices for Europe. For commodities, we use excess returns from the Bloomberg Commodity Index.

Besides data on asset returns, we obtain data for several control variables generally shown to have some forecasting power for asset returns, to be used in our predictive regressions. We control for the dividend yield of each equity market. In addition, across asset classes we control for country-specific term spread (the difference between the 10-year government bond and the three-month t-bill yields), the short-rate (three-month t-bill yield), U.S. default spread (the difference between Moody's BAA and AAA corporate bond spreads for U.S. issuers), and U.S. short rates (see for example Welch and Goyal (2008)), all obtained from Datastream. We also control for the 12-month time-series momentum of each underlying asset, as Moskowitz, Ooi, and Pedersen (2012) show strong predictive power of time-series momentum variables for equity, bond, and commodity returns, and as past returns may naturally correlate with past economic surprises. Lastly, we use our constructed global growth or inflation level and disagreement factors as control variables.

## 2.3 The behavior of macroeconomic surprises

In this section, we examine stylized facts in the behavior of macroeconomic surprises. We first show that macroeconomic surprises, as measured by the macroeconomic surprise nowcasts introduced in the previous section, do not follow a random walk, but rather exhibit positive autocorrelation. We call this phenomenon economic surprise momentum, which is observed locally in each region and globally, and is especially present in growth surprises. Furthermore, local surprises are strongly correlated with global surprises, indicating a common global component in economic surprises. Second, we examine forecasts and surprises on individual macroeconomic series and find that macroeconomic consensus forecasts are too rigid compared to actual changes. In other words, macroeconomic consensus forecasts underreact yielding autocorrelation in macroeconomic surprises. We show that economic surprise momentum occurs systematically due to (i) underreaction in economists' consensus forecasts to the series' time-series behavior *and*, novel to the literature, (ii) underreaction in consensus forecasts to information in surprises of other series.

### 2.3.1 Economic surprise momentum

We start our analysis by examining the behavior of the macroeconomic surprise nowcasts. These nowcasts give a real-time, aggregate measure of macroeconomic surprises across (almost) all macroeconomic releases. Figure 1 shows the 21-day moving average of the global growth surprise nowcast (black line) and global inflation surprise nowcast (grey line) over time. Global growth surprises behave pro-cyclical, dropping sharply for example during the Great Financial Crisis. Growth surprises are typically positive during periods of economic expansion and negative during recessions. This indicates that forecasts are typically too low during periods of expansions, but too high during periods of recessions. By contrast, the inflation surprise factor exhibit dynamics that are more erratic in variation and less aligned with periods of expansions and recessions. Second, inflation and growth surprises seem to be unrelated. This indicates that both factors capture different aspects of the economy. Third, remarkably, the surprise nowcasts seem to be positively autocorrelated. Positive surprises are typically followed-up by more positive surprises, while negative surprises are followed up by negative surprises. Next, we explore this potential autocorrelation structure in-depth.

Table 1 reports the descriptive statistics for the surprise nowcasts. The upper (lower) panel shows the local and global growth (inflation) nowcasts. Both growth and inflation surprises tend to have a slightly negative mean, indicating that forecasts are typically larger than the announcement value. This is reflected back in the proportion of positive (negative) observations. For example, almost 57% of the observations for the global growth surprise nowcasts are negative. The last three columns of table 1 shows the 1-month, 3-month and 12-month autocorrelations. We calculate these autocorrelations by using the 21-day sub-sampling method on the daily surprise nowcasts. Across all growth surprise nowcasts, we observe substantial 1-month autocorrelations, between 0.36 and 0.50. We also observe a positive, although weaker, autocorrelation between surprises at month $t$ and month $t-3$. 12-month autocorrelation, however, seem to be weaker and close to zero. These estimates suggest that economic surprises are correlated over time and non-random over a short time horizon. We observe this short-run economic surprise momentum across all regions, for both factors. For inflation surprises, the 1-month autocorrelation is closer (except for Japan), and equals 13% at the global level.

Next, to obtain a comprehensive overview of the autocorrelation structure of surprises we consider their corresponding autocorrelation function (ACF) using our 21-day resampling procedure. Figure 2 plots the results for the global growth and global inflation nowcasts. The upper (lower) plot provides the ACF for surprises in global growth (inflation). Consistent with our descriptive statistics, we observe positive and significant autocorrelations up to and including the sixth (monthly) lag for growth surprises. For inflation surprises only the one-month autocorrelation is positive and significant. Thus, growth surprises, in particular, exhibit a short-run momentum-like patterns: positive (negative) surprises tend to follow-up positive (negative) surprises, on average.

Next, we run several additional tests to confirm that the autocorrelation results are not mechanically driven by our forward filling exercise of economic surprises, nor our subsampling procedure. To reiterate, our 21-day resampling procedure allows us to calculate autocorrelations while controlling, to a large extent, for local persistency due forward filling. It could be that the forward filling of macroeconomic series at a lower frequency than monthly causes some spurious autocorrelation. To examine robustness, we repeat the above analysis only using the macroeconomic series that are released at a monthly frequency. Figure A.5 in the appendix shows the corresponding ACF of economic surprises. As before, we find growth surprises to be autocorrelated in the short run, witnessing sizable and significant autocorrelation up till six months. Surprises in inflation series remain are autocorrelated at the one month lag. In addition, we like to stress that the significant autocorrelation observed at the quarterly horizon cannot be explained by forward filling. Another possibility is that consensus forecasts are mechanically stale, for example due forecasters not updating forecasts in Bloomberg. Although we believe the latter to be rather unlikely, as performance of forecasters is well tracked, this could introduce spurious autocorrelation in surprises conditional upon autocorrelation in actuals. To examine, we also construct our surprise nowcasts by both excluding stale forecasts and using only monthly series. The corresponding autocorrelation plots, shown in figure A.6, reveal comparable results as before, with sizable and significant autocorrelations of especially growth surprises. We choose to not apply these filters as default as quarterly series included in our sample are generally important macroeconomic releases (e.g. GDP), and we believe stale forecasts to often reflect rational estimates of forecasters (for example for assumed random walk processes).

In addition to the surprise nowcast, we also constructed the level and disagreement nowcasts. Table 2 describes the correlation structure among the macroeconomic nowcasts. The lower triangle shows the correlations among global nowcasts, whereas the upper triangle shows the correlations among local nowcasts averaged across regions. We find that surprises are positively correlated with the level nowcast. For example, the correlation between the global growth surprise and level nowcast is 0.40, and 0.48 between the global inflation surprise and inflation level. This is in line with our earlier observation that surprises tend to follow the level factors. Growth surprises are high during economic expansions, when the actual growth nowcast also tends to be high. Surprises in inflation are negative, when actual inflation is low, which tends to coincide with periods of recession. Furthermore, global inflation is positively correlated with global growth. In addition, we observe a negative correlation between the level factors and the disagreement factors. Globally, the correlation is -0.74 within the growth category, and -0.12 within the inflation category. In times of economic expansion, disagreement among forecasters are low on average. Disagreement across categories tend to be positively correlated, whereas disagreement has a weak negative correlation with surprises.

### 2.3.2 What drives economic surprise momentum?

Next, we examine the behavior of forecasts and surprises across macroeconomic series using panel regressions. Empirically, several studies examine systematic patterns in macroeconomic forecasts using data of professional forecasters (Coibion and Gorodnichenko (2012, 2015); Broer and Kohlhas (2018); Bordalo et al. (2020)). These studies typically regress forecast errors on forecast revisions at the level of individual forecasters or consensus forecasts. As we have forecasts and actuals data available at the consensus level, we focus on explaining consensus forecast errors (i.e. surprises) with consensus forecasts and other public information available to forecasters.

First, we conduct a basic random walk test by regressing the surprise on a constant: $S_{i,t} = \alpha + \epsilon_t$. Under the null hypothesis of full information rational expectations, $\alpha$ will be equal to 0, since surprise on average should be equal to zero. We estimate the regression across the full breadth of our sample by pooling across all individual macroeconomic surprises within the growth or inflation category at their respective release dates (hence in an unbalanced panel data framework) and adjust the t-statistic for clustering in the time and variable ($i$) dimension. We conduct regressions for each region separately as well as the global pooled sample. Our sample includes between 4743 and 13585 observations across the four regions, and 34396 observation for the global sample. Note that we do not forward fill surprises, and as such do not need to resample at the 21-business day frequency to control for mechanically induced autocorrelation. In the first column of table 3, we report the estimated intercept and its corresponding t-statistic. We find, globally, that both growth and inflation surprises tend to be negative, however both not significantly different from zero. Under a random walk, we also expect surprise to be unpredictable. In table 3, column 2, we conduct a panel Breusch-Pagan test for autocorrelation in the residuals, and report the test statistic and p-value (within parenthesis). Globally, we find that residuals, obtained from the rationality test, are statistically significantly autocorrelated at the 1% (growth) to 5% (inflation) significance level, conforming the results reported in figure 2. Based on these findings, we can conclude that surprises behave differently from what we would except in a full information rational expectations framework.

Next, we test for systematic under- or over-reaction by estimating $S_t = \alpha + \beta * (F_t - A_{t-1}) + \epsilon_t$, in the spirit of Bondt and Thaler (1990)[16], and testing whether $\beta$ is significantly different from 0. The results reported in column (3) of table 3 show that $\beta$ is larger than zero in all instances, most significantly so at the global level. This holds for both growth and inflation surprises. Consequently, forecasts are too rigid, displaying underreaction. One argument behind this is that forecasters could potentially suffer from an anchoring heuristic, as shown by Campbell and Sharpe (2009). Following Campbell and Sharpe (2009), we next estimate $S_t = \alpha + \gamma_F * F_t + \gamma_A * A_h + \epsilon_t$. We use the previous announcement as an anchor ($A_h = A_{t-1}$). When $\gamma_A < 0$, consensus forecasts are biased towards lagged values of actual data releases (i.e. forecasts are 'anchored' towards recent release numbers). A positive coefficient estimate of $\gamma_F$ would imply that consensus forecasts systematically underreact over and above the anchoring on the previous announcement value. We report the estimates of $\gamma_F$ and $\gamma_A$ in columns (4) and (5), respectively. Globally, we find that growth variables have a significantly positive coefficient on $F_t$ and a significantly negative coefficient on $A_h$, hence showing general underreaction in consensus forecasts in combination with anchoring towards the previous announcement value. For inflation variables the coefficients have the same sign, but insignificantly so at the global level.

In addition, we conjecture that macroeconomic consensus forecasts also underreact to information in recent surprises in other macroeconomic data releases. To examine to what extent past surprises in other series $j$ explain the surprise in series $i$, we run the regression $S_{i,t} = \alpha + \gamma_1 * S_{index-i,t-1} + \epsilon_t$. Note that we skip surprises on other series released on the same day to avoid any potential overlap with news released at

[16] Bondt and Thaler (1990) estimate $A_t - A_{t-1} = \alpha + \gamma * (F_t - A_{t-1}) + \epsilon_t$ and test whether $\gamma$ is significantly different from 1. Note that $A_t - A_{t-1} = F_t - F_t + A_t - A_{t-1} = S_t + F_t - A_{t-1} = \alpha + \gamma * (F_t - A_{t-1}) + \epsilon_t$, thus $S_t = \alpha + (\gamma - 1) * (F_t - A_{t-1}) + \epsilon_t$. Hence, we can use $\beta = \gamma - 1$, and test whether $\beta$ significantly differs from 0.

$t$. A positive estimate of $\gamma_1$ implies that past economic surprises on other macroeconomic series are able to predict the current surprise in series $i$. In column (6) of table 3 we report the estimates. We find that $\gamma_1$ is always positive, most significantly so at the global level. This behavior is observed in both growth and inflation surprises. Hence, macroeconomic surprises are not only driven by underreaction to a series' own time series behavior, but also by underreaction in consensus forecasts to information in surprises of other series. Finally, the last three columns show the estimates of the slope coefficients in the regression $S_t = \alpha + \delta_1 * F_t + \delta_2 * A_{t-1} + \delta_3 * S_{index-i,t-1} + \epsilon_t$ , where we examine underreaction to a series own time series' behavior with 'cross-surprise' effects jointly. The results confirm results in earlier columns: surprises on a macroeconomic series are driven by underreaction in consensus forecasts to (i) a series' own time-series behavior, and (ii) information in past surprises of other macroeconomic series, simultaneously.

The above results of deviation from the full information rational expectations hypothesis in the direction of underreaction in consensus forecasts align with results reported in previous studies on macroeconomic forecasts. Most notably, Coibion and Gorodnichenko (2012, 2015) find evidence of underreaction in consensus forecasts of inflation and other macroeconomic variables, which they attribute to informational frictions such as rational inattention. Similarly, BGMS find evidence of underreaction in the aggregated consensus forecasts. There are, however, several important differences between these studies and ours. These studies examine the relationship between forecast errors (i.e. surprises) and revisions to forecasts of the same release over time, data we have not available, and typically study up to 22 series (mostly from the Survey of Professional Forecasters or Blue Chip Survey) at the quarterly frequency. Instead, we examine the relationship between surprises and consensus forecasts taken just before macroeconomic announcements, which is different from revisions to forecasts. In addition, we focus on high frequencies across a large panel of hundreds of macroeconomic forecasts across the globe, and aggregate these in an aggregate growth or inflation surprise metric.[17] Interestingly, underreaction in consensus forecasts is present in forecast revisions over time but also in forecasts just before macroeconomic announcements. Related, Campbell and Sharpe (2009) argue consensus forecasts on individual macroeconomic series systematically underreact due to an anchoring heuristic, which we confirm as yielding part of the observed underreaction. Further, and importantly, our results show consensus level underreaction is composed of two parts; underreaction to a series' time series behavior and underreaction to the information contained in recent surprises on other macroeconomic series. The latter is novel to the literature, and aligns both with the models of BGMS and Coibion and Gorodnichenko (2012, 2015) assuming forecasters have more (rational or irrational) inattention to other series. As such, we extend the find-

[17]Further, at the level of the individual forecaster (instead of the consensus level), BGMS and Broer and Kohlhas (2018) document overreaction in most of the macroeconomic forecasts. Interestingly, BGMS attribute these seemingly contradictory findings to overreaction to news at the individual forecaster level combined with information frictions, and show that their interaction determines consensus level over- or underreaction. As we lack data on individual forecasts we have to refrain from examining economic surprise momentum at the forecaster level.

ings of existing studies to information spillover across macroeconomic releases and to a large panel of forecasts across the globe taken just before macroeconomic releases.

To further examine the role of both drivers in economic surprises we decompose the autocovariance structure of the surprise nowcasts. To reiterate, the surprise nowcast ($S^r_{c,t}$) for each region $r$ by category $c$ is a weighted linear combination of all currently active underlying surprise series ($S^r_{i,c,t}$), where the weights are extracted via our recursive principal component analysis, as shown in equation 2.1. Since $S^r_{c,t}$ is a linear combination of its underlying surprise series, we can express the autocovariance of $S^r_{c,t}$ in terms of the autocovariance of $S^r_{i,c,t}$ and all cross-autocovariance terms. The autocovariance between $S^r_{c,t}$ and $S^r_{c,t-j}$ is equal to:

$$
\begin{aligned}
cov(S^r_{c,t}, S^r_{c,t-j}) = \sum_{i\in\omega_t} cov(\lambda^r_{i,c,t} S^r_{i,c,t}, \lambda^r_{i,c,t-j} S^r_{i,c,t-j}) + \\
\sum_{i\in\omega_t, i\neq k} \sum_{j\in\omega_t, k\neq i} cov(\lambda^r_{i,c,t} S^r_{i,c,t}, \lambda^r_{k,c,t-j} S^r_{k,c,t-j}) \quad (2.4)
\end{aligned}
$$

The first term denotes the contribution to the autocovariance in our surprise nowcasts from serial correlation *within* the individual time series. The second term denotes the contribution of cross-autocovariance *between* individual time series. One can think of the second part as a measure of spill-overs between surprises, indicating the extent to which lagged surprises on series $j$ affect future surprises on series $i$. For ease of interpretation, we scale all autocovariances by the variance of $S^r_{c,t}$, allowing for an interpretation in terms of autocorrelations.

We estimate equation 2.1 using daily data and our 21-business day subsampling technique for the first three monthly autocovariances. Figure 3 provides the results for the global growth and inflation categories. The results show positive autocorrelation on the nowcast-level for each lag. Decomposing the nowcast-level autocovariances reveals positive scaled autocovariances for individual surprise series, especially for growth surprises. This implies that individual surprise series, weighted by their factor loadings, exhibit short-run autocorrelation on a stand-alone basis, confirming the results reported in table 3. In addition, we document positive spillovers among surprise series for each category, on average. This again confirms that past surprises for a specific macroeconomic variable predict future surprises for other macroeconomic variables, i.e. surprises spillover across macroeconomic announcements. These terms, just like the nowcast-level autocorrelation, tend to decrease over the lag length. Our decomposition further reveals that the economic surprise momentum in growth variables originates for about 50% to 60% from autocovariance in individual surprise series, and 40% to 50% from cross-autocovariance across macroeconomic surprise series. By contrast, economic surprise momentum in inflation variables originates especially from cross-autocovariance across individual inflation surprise series. Overall, we conclude that short-run momentum in economic surprises is due to both momentum in surprises on individual macroeconomic series and spillover across surprise series.

## 2.4 Economic surprises and asset prices

We have shown that macroeconomic surprises do not behave randomly, but exhibit significant short-run positive autocorrelation. Moreover, macroeconomic forecasts display underreaction to a series' own time-series behavior *and* information contained in other releases, giving rise to economic surprise momentum. In this section, we examine whether investors anticipate this economic surprise momentum, or that economic surprises also predict returns. In an efficient market macroeconomic surprises should not systematically predict asset returns, with the exception of the incorporation of new information the minutes around the release. On the other hand, if the biases found in macroeconomic forecasts causing economic surprise momentum also impacts investment decisions, macroeconomic surprises would predict asset returns over the next days to months. To this end, we examine the predictability of returns across all major asset classes (equities, credits, commodities and government bond) across the major regions (U.S., U.K., the Eurozone and Japan). Note that our macroeconomic surprise metric does not include any market-based data, which could already reflect the market's interpretation of surprises.

### 2.4.1 Predictive regressions

To investigate the relationship between surprises and market returns, we use the following predictive regression model:

$$R_{t:t+h} = \alpha + \beta x_t + \epsilon_{t:t+h} \;\; \forall t = 1, \cdots, T-h \tag{2.5}$$

where $R_{t:t+h} = (R_{t+1} + 1) \times \cdots \times (R_{t+h} + 1) - 1$. $R_t$ denotes the excess return of an asset at day $t$. $x_t$, the variable of interest, denotes the (local or global, growth or inflation) surprise nowcast at day $t$. We use a monthly forecast horizon and sample frequency (hence avoiding overlapping observations) as starting point, as we have found the autocorrelation in economic surprises to be especially strong at this horizon. The one-but-last surprise value of the month is used to predict next month's return. We include a one-day implementation lag on the surprise value (hence not using the last surprise value of a month) to account for implementation frictions and potential overlap between predictor and predicted variables. We like to stress that this is a conservative choice which slightly weakens the evidence in favor of return predictability. Our null hypothesis equals $\beta = 0$, implying no predictive ability. Our alternative hypothesis equals $\beta \neq 0$, implying predictive ability. We estimate the regression by pooling observations across regions included in our sample and adjust the t-statistic for clustering in the time and asset dimension. The exception is commodities, for which we use regular OLS regressions with Newey-West corrected t-values.

Table 4 presents the predictive regression results for growth surprises. Several observations emerge. First, local growth surprises positively predict future equity market and credit market returns (columns 1), whereas it negatively, but insignificantly, predicts future bond market returns (column 5). Second, global growth surprises positively predict future equity, credit (columns 2) and commodities market returns (column 5), but not future bond returns. Splitting the local growth surprise up into its global and local-minus-global components reveals that the predictability predominantly stems from global growth surprises for equity and credit markets (columns 3). These effects remain robust to the inclusion of multiple control variables outlined in section 2 (columns 4 and 6), becoming also significantly negative for bond markets (column 8). The effects are generally also economically substantial. For example, in column 2, a one standard deviation (0.16) increase in the global growth surprise nowcast is associated with an average increase of 102 basis points in the subsequent monthly excess equity return.

In table 5 we presents the predictive regression results for inflation surprises. Inflation surprises positively predicts future equity market returns, an effect present using both local and global surprises. Further, global inflation surprises positively predict credit (albeit marginally) and commodity returns once control variables are included. In addition, inflation surprises negatively predict future bond returns once controlled for other factors. In terms of economic magnitude, a one standard deviation (0.34) increase in the global inflation surprise nowcast is associated with an average increase of 52 basis points in the subsequent monthly excess equity return. Overall, we can conclude that the results for inflation surprises are in direction comparable to the growth surprises results, although generally weaker in magnitude.

### 2.4.2 Investment strategies

The predictive regressions in section 2.4.1 have shown that macroeconomic surprises positively and significantly predicts future excess returns in risky asset classes. To further evaluate the economic significance for an investor, we next examine a simple real-time investment strategy that aims to exploit the predictive power. The investment strategy takes a position equal to to the 1-day lagged value of a growth or inflation nowcast on the last day of the month. This position is held for one month, after which it is updated using the subsequent end-of-the-month nowcast value. We pool the markets within an asset class into a global strategy by equally weighting the markets.

Table 6 presents the results. For each strategy we compute the annualized Sharpe ratio, and the CAPM alpha and beta relative to the corresponding underlying asset class or market. In panel A (B), we report the results for the global (local) surprises. For equity markets, we find that exploiting global or local growth surprises yields sizable and significant Sharpe ratios of 0.50 or 0.58, respectively. For comparison, the Sharpe ratio on the global equity market is 0.36 over our sample. In addition, we find significant positively CAPM alphas of 2.10% and 2.61% per year, respectively. These investment strategies also exhibit a relatively negative exposure towards the market, as reflected in its negative beta. Further, we find that Sharpe ratios are also sizable and significant for credits and commodities, although the CAPM alpha is not significant for credits. Results for bond markets are in sign similar to the predictive regression results, but are economically insignificant for local or global growth surprises. Investment strategies based on inflation surprises are generally weaker, yielding positive significant Sharpe ratios for equity markets (with values of 0.39 or 0.36 for global or local nowcasts, respectively), albeit with insignificant CAPM alphas. In bond markets results are only significant for Sharpe ratio of the local inflation surprise nowcasts strategy. Overall, our investment strategy results especially stresses out the economic significance of global growth surprises in predicting future returns of risky assets. Table A.5, in the appendix, shows that the results are generally robust at the individual asset level.

### 2.4.3 Varying the forecast horizon

The predictive regression results presented in section 2.4.1 consider a one-month forecasting horizon. In this section, we examine alternative forecasting horizons, ranging from one day up to six months. For daily, weekly, monthly, quarterly, or bi-annual forecast horizons we sample at the end of each day, Friday, month, quarter, or December/June cycle, respectively, hence using non-overlapping observations. We first present the results for growth surprise nowcasts in table 7. As before, Panel A (B) summarizes the results of predictive panel regressions using global (local) growth surprises. We find that the predictive power of global or local growth surprises for future equity, credit and commodity returns is generally present for forecasting horizons from one day up to three months, the horizons for which the autocorrelation in economic surprises is also the most prevalent (see figure 2). Increasing the forecasting horizon to six months causes most coefficients to become insignificant, especially once control variables are included. Further, bond markets returns are only significantly predictable using global surprises using a one month or three months forecasting horizon. Table 8, Panel A (B) presents the results across forecasting horizons for global (local) inflation surprise nowcasts. Akin to the results for growth surprises, global inflation surprises positively predict equity market returns for one day to three months forecast horizons, albeit most significantly so at the one month forecasting horizon. This aligns with economic surprise momentum being mainly present in inflation surprises on the one month horizon. Results change sign afterwards, with the global inflation surprise nowcast negatively predicting returns at the six months forecast horizon. Further, local inflation surprises significantly predict equity market returns until one month out. Predictability for credit and commodity market returns is generally present at the one month and three months forecasting horizon. On the other hand, global and local inflation surprises negatively predict future bond returns on forecast horizons between one day and three months, especially after including control variables. Overall, our results indicate that assets returns tend to be predictable using economic surprises during the next day to three months out.

### 2.4.4 Instantaneous announcement surprise versus revision surprise

In our definition a macroeconomic surprise consist of two parts: (1) the instantaneous announcement surprise, defined as the difference between the latest real-time announcement value and the economists' survey consensus, and (2) the revision surprise, which is the difference between the latest announced value and last period's announcement. The predictability of asset returns by economic surprises that we document can thus come from the first part, the second part, or both parts. Next, we decompose our surprise nowcasts into the instantaneous announcement surprise and the revision surprise parts and examine their roles in return predictability.

As before, we run predictive pooled regressions at the monthly frequency, but using the surprise nowcasts that either exclude the revision surprises, or the surprise nowcasts that only consists of the revisions surprises. Table 9, Panel A provides the pooled regression estimates using growth surprises excluding revisions for each asset class. We find that global growth surprise nowcasts excluding revisions positively predict future equity and credit returns, and negatively predicts government bond returns after including control variables. Predictive coefficients are still positive, but now insignificant for commodities. Further, as before predictability predominantly stems from global growth surprises, as local surprise nowcasts do not significantly exhibit predictive power on top of global surprise nowcasts, except for credit when including control variables. Table 9, Panel B provides the pooled regression estimates using only the revision surprises for each asset class. The results show that also revision surprises contain important information for asset prices, with significant predictability coefficients for all asset classes, including commodities. In addition, local surprise nowcasts now do significantly exhibit predictive power on top of global surprise nowcasts. In terms of relative importance we observe that for equities, credits and bonds the instantaneous announcement surprises yield the highest $R^2_{adj}$, while for commodities this holds for revision surprises. The result on revision surprises is, to our knowledge, novel in the literature: revision surprises in macroeconomic growth variables contain predictive power for future equity, credit, commodity, and bond returns. Revision in macroeconomic growth variables reflect the arrival of new information that is used to update the state of the economy, thereby affecting asset prices.

Table 10, Panel A (B) present the results for inflation surprises using instantaneous (revision) surprises. Results are more mixed, as we find that especially local instantaneous inflation surprises negatively predict the one-month ahead equity returns, whereas global revision surprises in inflation positively predict future equity returns. By contrast, the negative predictability of inflation surprises for bond returns mainly originates from the revision surprises. For credits and commodities we find no significant predictive power of instantaneous and revision surprises in inflation.

### 2.4.5 Do other weighting schemes matter?

One of the goals of this paper is to summarize a large cross-section of economic surprises into one real-time macroeconomic nowcast statistic over a wide time span using a sound and simple methodology. We have opted for using principal component analysis within pre-defined economic categories (i.e. employment, output, and sentiment series form our growth category, and inflation series our inflation category) using an recursive time window. Other methodologies exist that also allow for distilling a large cross-section of daily economic surprises into one measure. In this section, we consider multiple alternative econometric methodologies to construct aggregate macroeconomic surprise nowcasts to examine the robustness of return predictability by macroeconomic surprises.

The first, and most simple, alternative method is to use equal weights. As such,we construct growth and inflation surprise nowcasts by assigning equal weights to the corresponding individual surprise time series. Second, we construct an attention-based surprise nowcast by using each day the number of forecasters of a series $i$, scaled by the total number of forecaster-series combinations, as weights for series $i$. Macroeconomic variables that are widely followed and forecasted by economists may receive more attention from investors. Lastly, we use the three-pass regression filter (3PRF) introduced by Kelly and Pruitt (2015). PCA achieves dimension reduction by decomposing the predictor's covariance matrix into eigenvalues, thereby extracting predictive information according to the covariance within the cross-section. By contrast, in the 3PRF predictive information is extracted according to the covariance with the factors driving the dependent variable. The first pass of the 3PRF consists of time-series regressions where each predictor variable is the dependent variable and the proxy (the asset market return here) is the regressor. The second pass is a cross-sectional regression where the underlying surprise variables are used as predictors and the first-pass slope coefficients as regressors. The second pass slope coefficients are then used to predict market returns in a third pass predictive regression. This approach is especially advantageous when the set of predictors is large, as in our case. Kelly and Pruitt (2015) apply this methodology to forecast market returns and cash flow growth and find positive in-sample and out-of-sample $R^2$ that outperform OLS and PCA. We implement the 3PRF procedure on the individual surprise series using an expanding window (hence as before avoiding any look-ahead bias).[18]

We run predictive regressions using these alternative weighting schemes and compare this to our methodology (shown in the column 'PCA'). In table 11 we report the estimated slope coefficients and their t-values (between parenthesis). Considering the equal-weighted or attention-weighted surprise nowcasts, we find these to yield similar predictability for equity, bond, credit, and commodity market returns as using the PCA methodology. This holds for both growth surprises as inflation surprises. In fact, the predictive results are, on average, slightly stronger than the results obtained by our PCA methodology when measured by the t-values on the

[18] We use an initial training period of 24 months

predictive coefficients. Surprisingly, using the 3PRF methodology, which provides more weight to macroeconomic variables that correlate more with future asset returns, generally yields weaker results. Using the 3PRF methodology, only growth surprises significantly predict future equity market returns, while predictability is largely absent for the other asset classes. In table 12 we quantify the economic gains of trading on surprises based on the other alternative methodologies. Findings are in line with the predictive regression results. Irrespective of the implemented method, we find that exploiting growth surprises in equity markets yield positive and significant alphas varying between 0.55% and 4.48% per year, and Sharpe ratios between 0.45 and 0.65%. Similarly, exploiting growth surprises yield significant gains in credit and commodity markets, exploiting inflation surprises yield significant gains in equity markets, and results are generally most strong for the equal-weighted or attention-weighted approaches. In summary, we find that economic surprise momentum in asset returns is a phenomena that is robust across various nowcasting techniques.

## 2.5 Economic surprise momentum, return predictability and asset risk

The previous sections have shown that macroeconomic forecasts are predictably biased with short-run momentum in economic surprises and predictability in equity, credit, commodity, and bond market returns. In this section, we examine the link between the autocorrelation in macroeconomic surprises and return predictability and explore alternative sources of the documented predictability. First, we link the return predictability to the economic surprise momentum. Second, we examine whether economic surprises are related to increases in risk, as reflected in the future volatility or skewness of asset returns. Third, we test if surprises predict negative expected returns on risky assets.

### 2.5.1 Economic surprise momentum and return predictability

Returns on risky assets are predictable by macroeconomic surprises. If momentum in economic surprises drives return predictability, we would expect the return predictability to be stronger during periods where surprises continue in the same direction as lagged surprises (i.e. equal signs). To examine this, we estimate specification 2.5 separately for the cases in which the global surprise nowcasts have the same or different sign at the end of months $t$ and $t+1$.

The results for global surprises are presented in table 13. When current and future surprises are of the same sign, predictability is persistently present and stronger in terms of both t-value and R-squared for risky assets. For example, the global growth surprise positively and significantly predicts future equity returns when the sign is equal, with a coefficient of 7.10 and a t-value of 3.30. On the other hand, equity return predictability is absent when current and future global growth surprises have opposite signs (t-value = 0.88). We find similar results for credits, while predictability is slightly stronger for commodities in case of equal signs. Notably, predictability for bond returns becomes marginally significant in case of equal signs (t-value = -1.71). Panel B reports the results for global inflation surprises. Overall, we tend to find similar to higher R-squared values in case of equal signs with the exception of commodities. That said, patterns are less clear and consistent for inflation surprises, perhaps due to the overall weaker predictability of asset returns by inflation surprises. Table A.12 in the appendix shows similar results when considering local surprise nowcasts.

Next, we further split the sample based on the sign of individual autocorrelations or cross-autocorrelations, akin to our finding that short-run momentum in economic surprises is due to both momentum in surprises on individual macroeconomic series and spillover across surprise series. We proxy individual autocorrelations by the weighted sum of the contemporaneous surprise multiplied by its one-release lagged surprise, i.e. $\sum w_i w_{i,t-1} S_{i,t} S_{i,t-1}$. When this sum is positive, contemporaneous surprises have the same sign as their lag, and hence indicate continuation. We proxy cross-autocorrelations by the weighted sum of the contemporaneous surprise multiplied by all lagged surprises from other series, i.e. $\sum \sum_{i \neq j} w_i w_{j,t-1} S_{i,t} S_{j,t-1}$. When this term is positive, surprises in series $j$ are followed up by surprises with the same sign in the other series $i$. We re-run specification 2.5 separately for the cases in which either terms have the same or different signs at the end of month *t+1* relative to the end of month $t$.

The results for global growth surprises are shown in panel C of table 13. The predictive power of global growth surprises are higher when the sign of individual current and lagged surprises are equal instead of different. For equities, for example, we find a coefficient of 6.78 (t-value = 3.46) when surprises have equal own signs, whereas the coefficient is 3.75 and not significant (t-value = 0.98) when the own signs differ. We find similar results for credit or bond returns, while predictability for commodities is similar across the own and different sign sub-samples. Interestingly, we find that global growth surprises have significant predictive power especially when the cross signs among surprises are equal, including for commodities. Finally, panel D shows the results for global inflation surprises, indicating that a less consistent picture with the exception of commodities. Inflation surprise tend to contain predictive power for bond returns when the sign differ, both the own - and cross sign. Our results indicate return predictability from economic growth surprises (not inflation surprises) is driven by autocorrelation in macroeconomic surprises; positive (negative) surprises tend to be followed by future positive (negative) surprises yielding predictable asset returns, and these results originate in surprise momentum in the own-series and the spillover across series.

### 2.5.2 Return predictability and asset risks

Higher expected returns on risky assets after positive economic surprises may be the result of increases in risk or risk premia demanded by investors. To test this predictive link, we test whether economic surprises predicts realized volatility, defined as the sum of next month's daily squared returns. Table 14 reports the result for each asset class using growth surprises. We find no evidence of higher surprises leading to increased risks. For equity markets, we find that surprises in global economic growth variables negatively predicts future market volatility. However, we have seen that global growth variables positively predicts future equity market returns. These findings suggest that a premium for global growth surprises is not justified by changes in future volatility. We find similar patterns for credits, although insignificantly so after including control variables, and commodities. For bond markets, which over our sample correlate negatively with equity markets and hence could be seen as a hedge asset against bad states of the world, we find that global growth surprises negatively predicts both the future return and volatility.

Table 15 reports the estimates using inflation surprises. Results are generally in line with the results using growth surprises, albeit a magnitude weaker. Global inflation surprises negatively predicts future equity volatility, whereas it positively predicts future equity market returns. However, once controlled for other variables, the coefficient on inflation surprises becomes insignificant. Thus, the positive link between inflation surprises and equity returns do not coincide with decreased market volatility. Similarly, for other classes we also do not find that inflation surprises consistently predict lower future market volatility. As a robustness test, in appendix table A.6 and table A.7 we examine predicting future next month's realized skewness using economic surprises. These results indicate that asset return predictability does not coincide with predictability in future skewness across asset classes.

### 2.5.3 Return predictability and negative expected returns

Next, we test if economic surprises predict negative expected excess returns. A rational time-varying risk premia explanation is hard to align with negative expected returns on especially equities, but also credits. Rational finance models require that the aggregate stock market is a hedge against aggregate consumption risk in order to yield negative expected returns Baker and Wurgler (2000). Yet, expected equity premia can be negative if the predictability reflects a market inefficiency. We conjecture that a similar argument applies to credit markets, while no clear implication of negative expected returns can be drawn for bond or commodity markets.

Our findings generally reveal significantly negative expected equity and credit returns after the most negative surprises. The results reported in table 4 show a pooled univariate regression of equity returns on global growth surprise yields a constant estimate of 0.59 and growth surprise coefficient of 6.37, while the mean and standard deviation of the global growth surprise equal $-0.03$ and 0.16 (see table 1).

Hence, we observe negative forecasts for next month's equity premium for values of the global growth surprise of roughly half a standard deviation below its average. More precisely, the expected equity premium is forecasted to be negative in 28.9% (306 out of 1096) of cases. As estimates are imprecise, we next take forecast uncertainty into account by calculating the 90% confidence interval estimate around these predicted values. We predict negative expected excess equity returns in 20.4% of all observations, as shown in Panel A of table 16. The corresponding number for credit markets is 4.9%.

A possibility is that these results might be due to an omitted non-linearity in the relationship between the global growth surprise and the equity or credit premium. Potentially, predictability is only derived from medium to high values of the global growth surprise, while there might be no relationship between low values of the growth surprise and equity premia. This would imply that the results may still be aligned with a time-varying risk premium explanation. To allow for this possibility, we next calculate the cumulative prediction error of the above analysis over all negative predicted equity or credit premia months. The presence of convexity between global growth surprises and future return premia would result in a positive cumulative prediction error. We find a positive cumulative prediction error for equities equal to 55.1%. However, when we only focus on the 10.0% of predictions that are significantly negative, we find a cumulative prediction error of 30.9%. To formally control for convexity we follow Driesprong, Jacobsen, and Maat (2008), and estimate:

$$r_{t+1} = \alpha + \beta_1 S_t + \beta_2 D_{t+1} S_t + \epsilon_{t+1}$$

$D_{t+1}$ takes a value of one when predicted returns are negative (based on the estimates of the regression $R_{t+1} = \alpha + \beta S_t$), and zero otherwise. $\beta_1$ measures the predictive relationship between the global growth surprise and the next month's return for positive predicted returns. $\beta_2$ measures whether the relationship is different for negative expected return predictions. $\beta_1 + \beta_2$ measures the total predictive effect for negative expected return predictions. To test whether $\beta_1 + \beta_2$ is significantly different from zero, we use a Wald test.

The results are reported in Panel B of table 16. We find that the predictive effect is not different for negative or positive expected equity returns. The estimate of $\beta_1$ is equal to 4.57. If convexity is present, we expect $\beta_2$ to be equal to roughly $-\beta_1$. However, we find an estimate of $\beta_2$ equal to 3.14. In addition, we reject the hypothesis that $\beta_1 + \beta_2$ equals zero, implying a significant predictive relationship for the negative return domain. Again, the results for credit market returns are comparable.

The above two procedures depend on a relatively little number of observations in the negative expected return domain, and assume normality of the error term. Consequently, they might lack statistical power. As a final test, we therefore implement the minimum expected return test of Eleswarapu and Thompson (2007).

They propose to test the distance between the parameter estimates and the closest point in parameter space consistent with non-negativity, while employing a bootstrap procedure to deal with sample size issues and violations of normality. The results, reported in Panel C of table 16, show that the minimum expected equity (credit) market return of -3.12% (-0.85%) is significantly smaller than zero (p-value =0.00 or 0.01). Our findings for commodities markets are similar, while for bond markets we find zero predictions that are significantly negative and we reject the null hypothesis of no predictability of negative returns using the minimum expected return test. We report these latter results for sake of consistency, although we like to stress that the implication of negative expected returns is ambiguous for commodity and bond markets.

Furthermore, in table 17 we report the results using global inflation surprises. In contrast to growth surprises, inflation surprises predict less negative equity or credit returns. That said, in line with the growth results, we reject the hypothesis that $\beta_1 + \beta_2$ equals zero, implying a significant predictive relationship for the negative return domain. Lastly, the minimum expected return test indicates negative expected equity or credit market returns. In Appendix table A.8 and table A.9, we provide results using local surprise growth or inflation nowcasts, finding largely similar results. Overall, these tests indicate evidence of predictable negative returns on risky assets, a result that seems hard to reconcile with a market efficiency-based explanation. Our findings suggest that the predictability reflects a market inefficiency due to investors' expectations being sticky. By contrast, the results seem hard to reconcile with explanations based on risk, although we like to be cautious on such an interpretation as risk exposures, and especially risk premia, are not directly observable.

## 2.6 Conclusion

We comprehensively examine the behavior and impact of macroeconomic surprises across macroeconomic categories (i.e. growth and inflation), regions (i.e. U.S., U.K., Japan, and the Eurozone), and asset classes (i.e. equities, bonds, credits and commodities). As our first major contribution, we develop a novel real-time nowcast of macroeconomic surprises across hundreds of macroeconomic series that allows us to measure news comprehensively and in real-time. This surprise nowcast summarizes the wealth of macroeconomic releases into a single reading on growth or inflation surprises, which can be utilized to understand the aggregate behavior of macroeconomic surprises and its incorporation in asset prices.

As our second major contribution, we study biases in the behavior of macroeconomic surprises and macroeconomic forecasts across the globe, documenting a strong stylized fact. Macroeconomic surprises do not follow random walks but exhibit short-term positive autocorrelation. Positive (negative) surprises tend to be followed by positive (negative) surprises. This economic surprise momentum is especially strong in economic growth measures, and stems from autocorrelation in individual surprise series *and*, novel to the literature, cross-autocorrelation between surprise series. Consensus forecasts are predictably biased by underreacting to a series' own time series behavior *and* information contained in recent surprises on other macroeconomic series.

As our third major contribution, we examine return predictability originating from economic surprises. We argue that examining return predictability provides a natural test on rational versus irrational explanations of systematic patterns in economic surprises. Overall, growth surprises positively predicts returns on risky assets (i.e. equities, credits and commodities). Inflation surprises display a similar direction of predictability, albeit weaker. These findings are robust to control variables, are consistent across regions and alternative nowcast methods, and hold for predictability horizons up to about three months out. Novel to the literature, we also incorporate surprises due to revisions in macroeconomic data and show that also this revision component plays a substantial role in predicting returns. Further tests reveal that the return predictability is driven by the autocorrelation in macroeconomic surprises. Our results align with aggregate underreaction by macroeconomic forecasters and investors driven by expectational errors. Overall, we conclude that economic surprise momentum is a strong empirical phenomena in macroeconomic surprises and asset returns.

## 2.7 References

Ait-Sahalia, Y., Mykland, P. A., & Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, *18*(2), 351–416.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Vega, C. (2007). Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics*, *73*(2), 251–277.

Aruoba, S. B., Diebold, F. X., & Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, *27*(4), 417–427.

Baker, M., & Wurgler, J. (2000). The equity share in new issues and aggregate stock returns. *Journal of Finance*, *55*(5), 2219–2257.

Balduzzi, P., Elton, E. J., & Green, T. C. (2001). Economic news and bond prices: Evidence from the us treasury market. *Journal of Financial and Quantitative Analysis*, *36*(4), 523–543.

Baltussen, G., & Soebhag, A. (2022). Caught by surprise: how markets respond to macroeconomic news. *Available at SSRN 4059314*.

Baltussen, G., van Bekkum, S., & Da, Z. (2019). Indexing and stock market serial dependence around the world. *Journal of Financial Economics*, *132*(1), 26–48.

Beber, A., & Brandt, M. W. (2009). When it cannot get better or worse: The asymmetric impact of good and bad news on bond returns in expansions and recessions. *Review of Finance*, *14*(1), 119–155.

Beber, A., Brandt, M. W., & Luisi, M. (2015). Distilling the macroeconomic news flow. *Journal of Financial Economics*, *117*(3), 489–507.

Beshears, J., Choi, J. J., Fuster, A., Laibson, D., & Madrian, B. C. (2013). What goes up must come down? experimental evidence on intuitive forecasting. *American Economic Review*, *103*(3), 570–74.

Bondt, W. F. M. D., & Thaler, R. H. (1990). Do security analysts overreact? *The American Economic Review*, *80*(2), 52–57.

Bordalo, P., Gennaioli, N., Ma, Y., & Shleifer, A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, *110*(9), 2748–82.

Bordalo, P., Gennaoli, N., Porta, R. L., & Shleifer, A. (2019). Diagnostic expectations and stock returns. *Journal of Finance*, *74*(6), 2839-2874.

Bouchaud, J.-P., Krüger, P., Landier, A., & Thesmar, D. (2019). Sticky expectations and the profitability anomaly. *Journal of Finance*, *74*(2), 639-674.

Boyd, J. H., Hu, J., & Jagannathan, R. (2005). The stock market's reaction to unemployment news: Why bad news is usually good for stocks. *Journal of Finance*, *60*(2), 649–672.

Broer, T., & Kohlhas, A. (2018). Forecaster (mis-) behavior. *Working paper*.

Campbell, & Diebold, F. (2009). Stock returns and expected business conditions: Half a century of direct evidence. *Journal of Business & Economic Statistics*, *27*(2), 266–278.

Campbell, & Sharpe, S. (2009). Anchoring bias in consensus forecasts and its effect on market prices. *Journal of Financial and Quantitative Analysis*, *44*(2), 369–390.

Chousakos, K., & Giamouridis, D. (2020). Harvesting macroeconomic risk premia. *The Journal of Portfolio Management*, *46*(6), 1-16.

Coibion, O., & Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, *120*(1), 116–159.

Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, *105*(8), 2644–78.

Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of Economic Literature*, *49*(1), 72–100.

Croushore, D., & Stark, T. (2003). A real-time data set for macroeconomists: Does the data vintage matter? *Review of Economics and Statistics*, *85*(3), 605–617.

Driesprong, G., Jacobsen, B., & Maat, B. (2008). Striking oil: Another puzzle? *Journal of Financial Economics*, *89*(2), 307-327.

Eleswarapu, V., & Thompson, R. (2007, 02). Testing for negative expected market return premia. *Journal of Banking & Finance*, *31*, 1755-1770.

Fama, E. F., Fisher, L., Jensen, M. C., & Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, *10*(1), 1–21.

Faust, J., Rogers, J. H., Wang, S.-Y. B., & Wright, J. H. (2007). The high-frequency response of exchange rates and interest rates to macroeconomic announcements. *Journal of Monetary Economics*, *54*(4), 1051–1068.

Flannery, M. J., & Protopapadakis, A. A. (2002). Macroeconomic factors do influence aggregate stock returns. *Review of Financial Studies*, *15*(3), 751–782.

Ghysels, E., Horan, C., & Moench, E. (2014). Forecasting through the rearview mirror: Data revisions and bond return predictability. *Review of Financial Studies*.

Gilbert, T. (2011). Information aggregation around macroeconomic announcements: Revisions matter. *Journal of Financial Economics*, *101*(1), 114–131.

Gilbert, T., Scotti, C., Strasser, G., & Vega, C. (2017). Is the intrinsic value of a macroeconomic news announcement related to its asset price impact? *Journal of Monetary Economics*, *92*, 78–95.

Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *Review of Financial Studies*, *20*(4), 1113–1138.

Kelly, B., & Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, *186*(2), 294–316.

Krueger, A. B., & Fortson, K. N. (2003). Do markets respond more to more reliable labor market data? a test of market rationality. *Journal of the European Economic Association*, *1*(4), 931–957.

Landier, A., Ma, Y., & Thesmar, D. (2019). Biases in expectations: Experimental evidence. *Working paper, available at SSRN 3046955*.

La Porta, R. (1996). Expectations and the cross-section of stock returns. *Journal of Finance*, *51*(5), 1715–1742.

Ludvigson, S. C., & Ng, S. (2009). Macro factors in bond risk premia. *Review of Financial Studies*, *22*(12), 5027–5067.

McQueen, G., & Roley, V. V. (1993). Stock prices, news, and business conditions. *Review of Financial Studies*, *6*(3), 683–707.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, *41*(5), 867–887.

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, *104*(2), 228–250.

Paye, B. S. (2012). 'déjà vol': Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*, *106*(3), 527–546.

Pedersen, L. H. (2019). *Efficiently inefficient: how smart money invests and market prices are determined*. Princeton University Press.

Scotti, C. (2016). Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises. *Journal of Monetary Economics*, *82*, 1–19.

Smirlock, M. (1986). Inflation announcements and financial market reaction: Evidence from the long-term bond market. *Review of Economics and Statistics*, 329–333.

Stambaugh, R. F. (1997). Analyzing investments whose histories differ in length. *Journal of Financial Economics*, *45*(3), 285–331.

Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, *4*, 351–394.

Urich, T., & Wachtel, P. (1984). The effects of inflation and money supply announcements on interest rates. *Journal of Finance*, *39*(4), 1177–1188.

Vrugt, E. B. (2009). U.s. and japanese macroeconomic news and stock market volatility in asia-pacific. *Pacific-Basin Finance Journal*, *17*(5), 611–627.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*(4), 1455–1508.

## 2.8 Figures & Tables

Figure 1: **Global growth and inflation surprises nowcast.** The figure plots the 21-day moving average of the global growth surprise nowcast (black line, left y-axis) and global inflation nowcast (grey line, right y-axis) over time. Regional surprise nowcasts are constructed according to equation (1), which are subsequently equally-weighted into a global surprise nowcast. The sample period runs from 31-03-1997 until 31-12-2019.

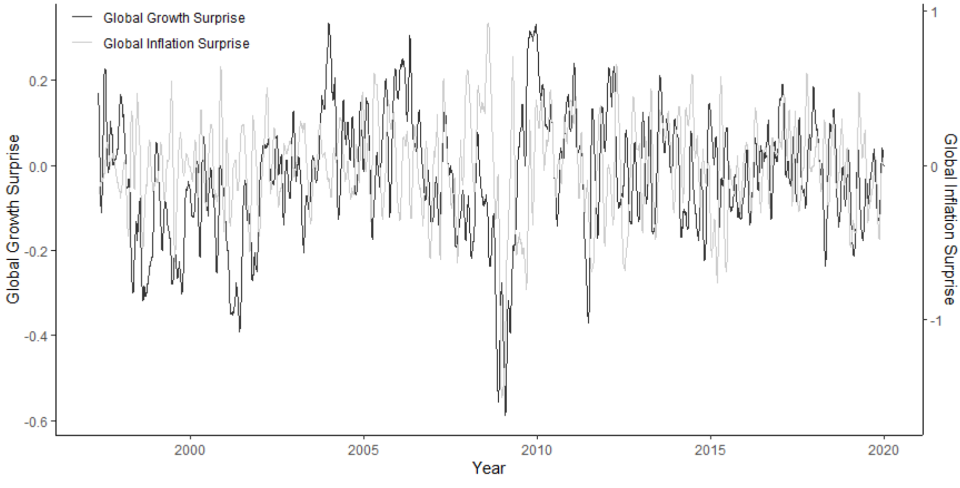


Figure 2: **Economic surprise momentum.** The upper (lower) plot depicts the autocorrelation pattern of the global growth (inflation) surprise nowcast. Autocorrelations are calculated at the monthly frequency (using 21 business-day sub-sampling) for lags 1 to 36. Dotted lines indicate 5% significance levels. The sample runs from 31-03-1997 until 31-12-2019.

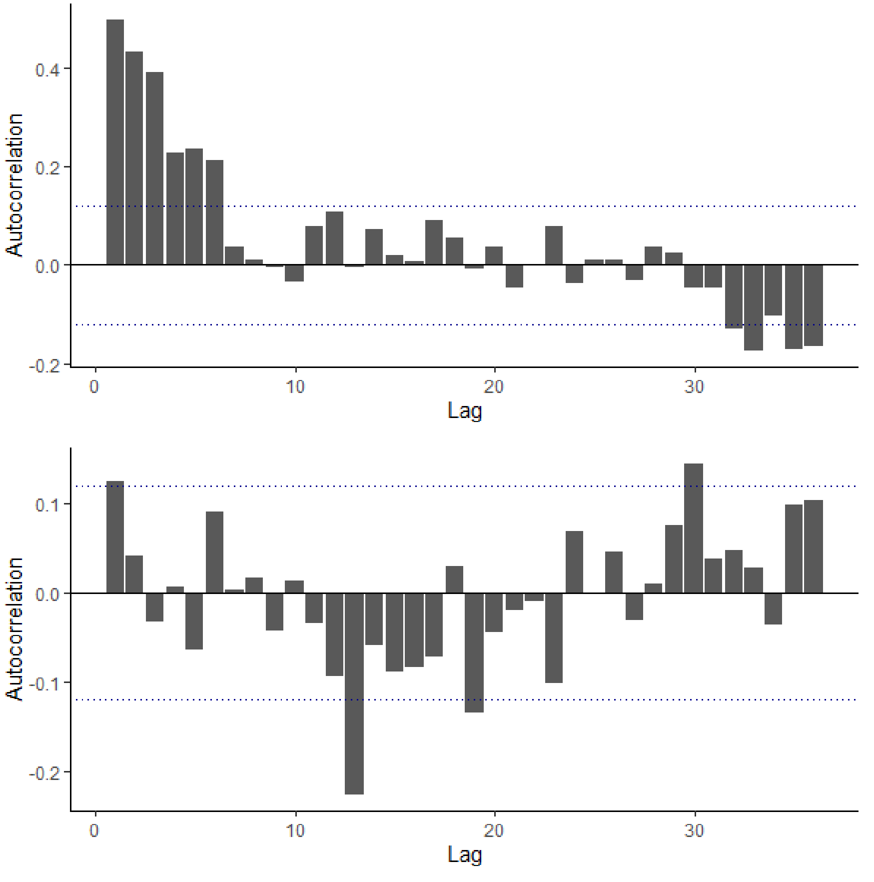


Figure 3: **Decomposition of economic surprise momentum.** This figure shows the autocovariance decomposition (equation 4) for the global growth (upper plot) and inflation (lower plot) nowcasts. The white bars indicate the scaled autocovariance at lag $j$ of the surprise nowcast for a given category. The black bars show the weighted average of the scaled autocovariances of all underlying macroeconomic surprise numbers. Lastly, the grey bars show the cross-autocovariances between all underlying macroeconomic surprise series. Results are calculated at the monthly frequency (using 21 business-day sub-sampling) for lags 1 to 3. We scale all numbers by the full sample variance of the corresponding surprise nowcast to allow for an autocorrelation interpretation. The sample period runs from 31-03-1997 until 31-12-2019.

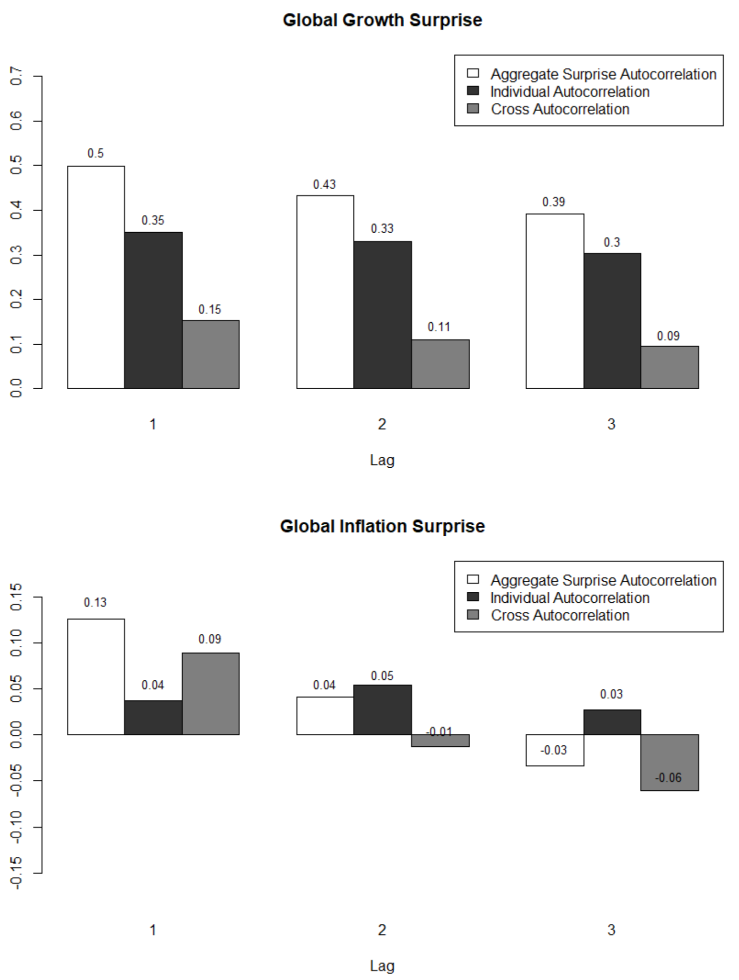


Table 1: **Descriptive statistics of global macroeconomic surprise nowcasts.** This table shows summary statistics of the macroeconomic surprise nowcasts per region and globally. Regional surprise nowcasts (for US, UK, JP and EU) are constructed using equation (1). Global surprise nowcasts (GL) are the equal-weighted average of the regional nowcasts. Shown are the number of observations (obs.), time-series mean values (mean), standard deviation (sd), minimum (min), maximum (max), the proportion of positive observations (pos), and the proportion of negative observations (neg) of growth ($S_g$) and inflation ($S_i$) nowcasts. $\hat{\rho}_i$ denotes the estimated autocorrelation between month $t$ and $t-i$ using 21 business day subsampling. The nowcasts have a daily frequency starting from 31-03-1997 until 31-12-2019.

| Var. | obs. | mean | sd | pos | neg | $\hat{\rho}_1$ | $\hat{\rho}_3$ | $\hat{\rho}_{12}$ |
|---|---|---|---|---|---|---|---|---|
| $S_{g,US}$ | 5,936 | -0.02 | 0.29 | 0.49 | 0.51 | 0.36 | 0.28 | 0.07 |
| $S_{g,UK}$ | 5,866 | -0.14 | 0.30 | 0.31 | 0.69 | 0.44 | 0.37 | 0.39 |
| $S_{g,JP}$ | 5,216 | 0.03 | 0.30 | 0.57 | 0.43 | 0.43 | 0.37 | 0.03 |
| $S_{g,EU}$ | 5,898 | 0.03 | 0.20 | 0.54 | 0.46 | 0.37 | 0.35 | -0.00 |
| $S_{g,GL}$ | 5,936 | -0.03 | 0.16 | 0.43 | 0.57 | 0.50 | 0.39 | 0.11 |
| $S_{i,US}$ | 5,903 | 0.03 | 0.45 | 0.54 | 0.45 | 0.06 | -0.02 | -0.08 |
| $S_{i,UK}$ | 5,883 | -0.01 | 0.68 | 0.49 | 0.51 | -0.20 | 0.01 | 0.10 |
| $S_{i,JP}$ | 4,698 | -0.06 | 0.64 | 0.47 | 0.53 | 0.59 | 0.27 | -0.35 |
| $S_{i,EU}$ | 5,905 | 0.00 | 0.49 | 0.48 | 0.52 | -0.09 | -0.16 | 0.36 |
| $S_{i,GL}$ | 5,905 | -0.01 | 0.34 | 0.51 | 0.49 | 0.13 | -0.03 | -0.09 |

Table 2: **Macroeconomic nowcast correlations.** The table reports the correlations of the various macroeconomic growth and inflation nowcasts. $A$, $S$ and $FD$ denote the actual, surprise, and disagreement nowcasts, respectively. Subscript $g$ ($i$) denotes growth (inflation). The lower triangle below the diagonal shows the correlations among the global nowcasts. The upper triangle above the diagonal shows the correlation among local factors, averaged across regions. The sample is at a daily frequency and runs from 31-03-1997 until 31-12-2019. Within parenthesis we provide the p-value of the correlation. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| | $A_g$ | $FD_g$ | $S_g$ | $A_i$ | $FD_i$ | $S_i$ |
|---|---|---|---|---|---|---|
| $A_g$ | | -0.51*** | 0.32*** | 0.27*** | -0.25*** | 0.14*** |
| | | (0.00) | (0.00) | (0.00) | (0.00) | (0.02) |
| $FD_g$ | -0.74*** | | -0.15*** | -0.09* | 0.24*** | -0.10*** |
| | (0.00) | | (0.00) | (0.08) | (0.00) | (0.01) |
| $S_g$ | 0.40*** | -0.22*** | | 0.05*** | -0.02 | 0.13*** |
| | (0.00) | (0.00) | | (0.00) | (0.23) | (0.25) |
| $A_i$ | 0.26*** | -0.15*** | -0.03** | | -0.11*** | 0.49*** |
| | (0.00) | (0.00) | (0.03) | | (0.00) | (0.00) |
| $FD_i$ | -0.52*** | 0.45*** | -0.09*** | -0.12*** | | -0.08** |
| | (0.00) | (0.00) | (0.00) | (0.00) | | (0.02) |
| $S_i$ | 0.19*** | -0.14*** | 0.12*** | 0.48*** | -0.17*** | |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | |

Table 3: **Explaining macroeconomic surprises.** The table summarizes the results of panel regressions across all macroeconomic growth (Panel A) or inflation (Panel B) series, release-by-release. The first column of both panels report the estimated constant, $\hat{\alpha}$, from the regression $S_{i,t} = \alpha + \epsilon_t$, with $S_{i,t}$ scaled by the standard deviation of the surprise $S_{i,t}$, and its Breusch-Godfrey test statistic (BG) in the second column. The next column shows $\hat{\beta}$ from the regression $S_{i,t} = \alpha + \beta * (F_{i,t} - A_{i,t-1}) + \epsilon_{i,t}$. Third, we estimate $S_{i,t} = \alpha + \gamma_F * F_{i,t} + \gamma_A * A_{i,t-1} + \epsilon_{i,t}$, and report $\hat{\gamma_F}$ and $\hat{\gamma_A}$, respectively. Fourth, we estimate $S_{i,t} = \alpha + \gamma_1 * S_{index-i,t-1} + \epsilon_{i,t}$, and report their estimates. Lastly, we report the slope estimates of the regression $S_{i,t} = \alpha + \delta_1 * F_{i,t} + \delta_2 * A_{i,t-1} + \delta_3 * S_{index-i,t-1} + \epsilon_{i,t}$. Below all estimates, we provide the corresponding t-values within parenthesis. For the BG statistics, the $p$-values are shown between parenthesis. All regressions are pooled, and standard errors are corrected for clustering in the time and macroeconomic series dimensions. The sample period runs from 31-03-1997 until 31-12-2019. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| | **Panel A: Growth surprises** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\alpha}$ | BG | $\hat{\beta}$ | $\hat{\gamma_F}$ | $\hat{\gamma_A}$ | $\hat{\gamma_1}$ | $\hat{\delta_1}$ | $\hat{\delta_2}$ | $\hat{\delta_3}$ |
| US | 0.00 | 87.24*** | 0.12*** | 0.08*** | -0.07*** | 0.74*** | 0.08*** | -0.07*** | 0.65*** |
| | (0.25) | (0.00) | (3.97) | (2.83) | (-2.86) | (2.72) | (2.90) | (-2.93) | (2.53) |
| UK | -0.02 | 19.79*** | 0.01 | 0.03 | -0.06 | 0.04 | 0.04 | -0.07* | 0.06 |
| | (-0.73) | (0.00) | (0.28) | (0.67) | (-1.39) | (0.10) | (1.01) | (-1.74) | (0.16) |
| JP | -0.03 | 11.53*** | 0.04 | 0.01 | -0.03 | 0.37 | 0.00 | -0.02 | 0.37 |
| | (-1.26) | (0.01) | (1.08) | (0.20) | (-0.94) | (1.07) | (0.04) | (-0.77) | (1.07) |
| EU | -0.03 | 77.40*** | 0.14*** | 0.10*** | -0.09*** | 0.73*** | 0.09*** | -0.08*** | 0.64** |
| | (-1.17) | (0.00) | (4.52) | (3.97) | (-3.69) | (2.66) | (3.24) | (-3.07) | (2.31) |
| Global | -0.01 | 140.58*** | 0.09*** | 0.06*** | -0.06*** | 0.59*** | 0.06*** | -0.05*** | 0.53*** |
| | (-1.27) | (0.00) | (4.59) | (3.59) | (-3.62) | (3.40) | (3.27) | (-3.30) | (3.19) |
| | **Panel B: Inflation surprises** | | | | | | | | |
| | $\hat{\alpha}$ | BG | $\hat{\beta}$ | $\hat{\gamma_F}$ | $\hat{\gamma_A}$ | $\hat{\gamma_1}$ | $\hat{\delta_1}$ | $\hat{\delta_2}$ | $\hat{\delta_3}$ |
| US | -0.04 | 8.56*** | 0.04 | 0.00 | 0.01 | 0.34 | -0.01 | 0.03 | 0.38 |
| | (-1.15) | (0.04) | (0.86) | (0.03) | (0.34) | (0.93) | (-0.17) | (0.71) | (1.10) |
| UK | 0.08 | 15.99*** | 0.03 | 0.01 | -0.02 | 1.06* | -0.02 | 0.01 | 1.10* |
| | (1.44) | (0.00) | (0.79) | (0.06) | (-0.25) | (1.70) | (-0.20) | (0.09) | (1.67) |
| JP | 0.04 | 0.65*** | 0.07* | 0.22 | -0.24 | 0.35 | 0.18 | -0.20 | 0.28 |
| | (1.13) | (0.89) | (1.87) | (1.38) | (-1.62) | (1.17) | (1.22) | (-1.43) | (0.89) |
| EU | -0.04 | 16.81*** | 0.05 | 0.11*** | -0.00 | 0.68*** | 0.12*** | 0.00 | 0.58*** |
| | (-1.35) | (0.00) | (1.49) | (3.63) | (-0.03) | (3.58) | (4.78) | (0.02) | (4.61) |
| Global | -0.00 | 9.99*** | 0.05** | 0.03 | -0.02 | 0.56*** | 0.03 | -0.01 | 0.53*** |
| | (-0.19) | (0.02) | (2.44) | (1.09) | (-0.82) | (2.76) | (0.94) | (-0.54) | (2.77) |

Table 4: **Growth surprises and asset returns.** The table summarizes the results of predictive panel regressions. We regress one-month ahead excess returns on end-of-month growth surprises for different asset classes (equity markets (Equity), 10-year government bonds (Bonds), credit indices (Credits), and the BCOM commodity index (Commodities)). We apply a one day implementation lag for predicting future returns. Shown are the predictive regression estimates, its corresponding t-values, the number of observations (Obs.), and the adjusted $R^2$. Subscript $l(g)$ indicates the local (global) macroeconomic surprise nowcast, while $l-g$ indicates the local-minus-global surprise nowcast. In applicable columns we control for the global level and disagreement growth nowcast, 12-months time-series momentum, the term spread, the risk-free rate, and the U.S. default spread across asset classes. For equity markets we also control for the dividend-yield. The sample period runs from 31-03-1997 until 31-12-2019. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| | Equity | | | | Bonds | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $S_l$ | 2.25*** | | | | −0.04 | | | |
| | (3.11) | | | | (−0.70) | | | |
| $S_g$ | | 6.37*** | 6.47*** | 5.32*** | | −0.10 | −0.11 | −0.26** |
| | | (3.42) | (3.56) | (3.17) | | (−1.27) | (−1.36) | (−2.42) |
| $S_{l-g}$ | | | 0.54 | 0.56* | | | −0.01 | −0.01 |
| | | | (1.41) | (1.76) | | | (−0.09) | (−0.14) |
| $C$ | 0.45** | 0.59*** | 0.57*** | 1.78 | 0.03*** | 0.03*** | 0.03*** | −0.13 |
| | (2.04) | (2.83) | (2.69) | (1.64) | (3.54) | (3.66) | (3.33) | (−0.53) |
| Controls | NO | NO | NO | YES | NO | NO | NO | YES |
| Obs. | 1,056 | 1,096 | 1,056 | 1,044 | 1,056 | 1,096 | 1,056 | 1,044 |
| $R^2_{adj}$ | 1.5% | 3.6% | 3.7% | 4.1% | 0.1% | 0.4% | 0.4% | 3.0% |

| | Credits | | | | Commodities | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $S_l$ | 0.95** | | | | | |
| | (2.23) | | | | | |
| $S_g$ | | 1.96* | 1.95** | 2.33** | 6.94*** | 6.75*** |
| | | (1.93) | (1.98) | (2.13) | (2.78) | (3.08) |
| $S_{l-g}$ | | | 0.44** | 0.50** | | |
| | | | (2.07) | (2.32) | | |
| C | 0.25** | 0.27** | 0.26** | 0.03 | 0.05 | −1.21 |
| | (1.96) | (2.05) | (2.02) | (0.06) | (0.20) | (-0.66) |
| Controls | NO | NO | NO | YES | NO | YES |
| Obs. | 824 | 824 | 824 | 780 | 274 | 268 |
| $R^2_{adj}$ | 1.9% | 2.7% | 2.9% | 3.9% | 5.3% | 4.0% |

Table 5: **Inflation surprises and asset returns.** The table summarizes the results of predictive panel regressions. We regress one-month ahead excess returns on end-of-month inflation surprises for different asset classes (equity markets (Equity), 10-year government bonds (Bonds), credit indices (Credits), and the BCOM commodity index (Commodities)). We apply a one day implementation lag for predicting future returns. Shown are the predictive regression estimates, its corresponding t-values, the number of observations (Obs.), and the adjusted $R^2$. Subscript $l(g)$ indicates the local (global) macroeconomic surprise nowcast, while $l-g$ indicates the local-minus-global surprise nowcast. In applicable columns we control for the global level and disagreement inflation nowcast, 12-months time-series momentum, the term spread, the risk-free rate, and the U.S. default spread across asset classes. For equity markets we also control for the dividend-yield. The sample period runs from 31-03-1997 until 31-12-2019. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| | Equity | | | | Bonds | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $S_l$ | 0.68*** | | | | −0.03** | | | |
| | (2.94) | | | | (−2.28) | | | |
| $S_g$ | | 1.53** | 1.42** | 2.49*** | | −0.05 | −0.05 | −0.08** |
| | | (2.35) | (2.17) | (3.08) | | (−1.60) | (−1.52) | (−2.06) |
| $S_{l-g}$ | | | 0.20*** | 0.18** | | | −0.01*** | −0.01 |
| | | | (3.99) | (2.32) | | | (−3.86) | (−0.82) |
| C | 0.43* | 0.39* | 0.43* | 1.22 | 0.03*** | 0.03*** | 0.03*** | −0.04 |
| | (1.82) | (1.73) | (1.83) | (0.95) | (3.33) | (3.84) | (3.37) | (−0.88) |
| Controls | NO | NO | NO | YES | NO | NO | NO | YES |
| Obs. | 1,032 | 1,088 | 1,032 | 982 | 1,032 | 1,088 | 1,032 | 982 |
| $R^2_{adj}$ | 0.5% | 1.1% | 0.9% | 5.9% | 0.5% | 0.8% | 0.7% | 4.1% |
| | Credits | | | | Commodities | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | | |
| $S_l$ | 0.01 | | | | | | | |
| | (0.05) | | | | | | | |
| $S_g$ | | 0.42 | 0.36 | 0.99* | 1.31 | 2.25** | | |
| | | (1.55) | (1.38) | (1.95) | (1.32) | (2.46) | | |
| $S_{l-g}$ | | | −0.29 | −0.25 | | | | |
| | | | (−1.52) | (−1.41) | | | | |
| C | 0.26** | 0.26** | 0.27** | 0.38 | −0.14 | -0.39 | | |
| | (2.01) | (2.02) | (2.02) | (0.78) | (−0.46) | (-0.28) | | |
| Controls | NO | NO | NO | YES | NO | YES | | |
| Obs. | 824 | 824 | 824 | 780 | 272 | 255 | | |
| $R^2_{adj}$ | −0.1% | 0.7% | 1.1% | 6.9% | 0.7% | 3.8% | | |

Table 6: **Macroeconomic surprise strategy.** This table summarizes the results of macroeconomic surprise investment strategies. At the end of each month long (short) positions are taken based on the sign of the macroeconomic growth or inflation surprise nowcast. Strategies use a one day implementation lag and hold positions for one month. Panel A (Panel B) reports the results for strategies based on global (local) macroeconomic surprises. We report the Sharpe ratio, the annualized $\hat{\alpha}$ in %, and the market exposure ($\hat{\beta}$) of each investment strategy relative to the corresponding global asset class returns from regressing $R_{s,t} = \alpha + \beta R_{m,t}$. The sample runs from 31-03-1997 until 31-12-2019 and consists of monthly non-overlapping observations. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| **Panel A: Global surprises** | Growth Surprise | | | Inflation Surprise | | |
|---|---|---|---|---|---|---|
| | Sharpe | $\hat{\alpha}$ | $\hat{\beta}$ | Sharpe | $\hat{\alpha}$ | $\hat{\beta}$ |
| Equity | 0.50** | 2.10** | -0.09** | 0.39* | 2.48 | -0.05 |
| | (2.40) | (2.35) | (-2.37) | (1.86) | (1.56) | (-0.94) |
| Bonds | -0.28 | -0.00 | -0.08 | -0.31 | -0.06 | -0.08 |
| | (-1.33) | (-0.24) | (-1.56) | (-1.46) | (-1.15) | (-0.88) |
| Credits | 0.44* | 0.56 | -0.04 | 0.37 | 0.69 | 0.01 |
| | (1.86) | (1.37) | (-0.80) | (1.58) | (1.38) | (0.20) |
| Commodities | 0.64*** | 1.85* | -0.08** | 0.32 | 2.00 | -0.02 |
| | (3.03) | (2.37) | (-1.99) | (1.51) | (1.21) | (-0.36) |
| **Panel B: Local surprises** | Growth Surprise | | | Inflation Surprise | | |
| | Sharpe | $\hat{\alpha}$ | $\hat{\beta}$ | Sharpe | $\hat{\alpha}$ | $\hat{\beta}$ |
| Equity | 0.58*** | 2.61** | -0.10** | 0.36** | 2.55 | -0.07 |
| | (3.91) | (2.58) | (-2.28) | (2.42) | (1.48) | (-1.20) |
| Bonds | -0.18 | 0.01 | -0.12 | -0.30** | -0.05 | -0.12 |
| | (-1.24) | (0.54) | (-1.69) | (-2.00) | (-0.97) | (-1.04) |
| Credits | 0.44*** | 0.69 | -0.05 | 0.03 | 0.22 | -0.06 |
| | (2.95) | (1.65) | (-0.74) | (0.20) | (0.44) | (-0.72) |

Table 7: **Growth surprises and asset returns: different forecasting horizons.** The table summarizes the results of predictive panel regressions for different forecasting horizons. Regression specifications follow table 4. In Panel A we use global growth surprises, while in Panel B we use local growth surprises. The sample period runs from 31-03-1997 until 31-12-2019 and consists of non-overlapping observations. For weekly, monthly, quarterly, or bi-annual forecast horizons we sample at the end of each friday, month, quarter, or December/June cycle, respectively. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| **Panel A: Global** | Equity | | Bonds | | Credits | | Commodities | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 1 day | 0.17* | 0.18* | -0.001 | -0.00 | 0.07* | 0.10** | 0.20** | 0.25** |
| | (1.81) | (1.93) | (-0.33) | (-0.89) | (1.71) | (1.99) | (2.25) | (2.39) |
| 1 week | 1.02** | 1.09** | -0.00 | -0.01 | 0.37* | 0.55** | 1.14** | 1.41*** |
| | (1.98) | (2.24) | (-0.57) | (-1.08) | (1.73) | (2.12) | (2.60) | (2.81) |
| 1 month | 6.37*** | 5.32*** | -0.10 | -0.16** | 1.96* | 2.33** | 6.94*** | 6.75*** |
| | (3.42) | (3.17) | (-1.27) | (-2.00) | (1.93) | (2.12) | (2.78) | (3.08) |
| 3 months | 11.74** | 9.58* | -0.14 | -0.46* | 5.83* | 7.80** | 16.02* | 21.39** |
| | (2.19) | (1.66) | (-0.65) | (-1.79) | (1.86) | (2.10) | (1.95) | (2.33) |
| 6 months | 15.07* | 12.93 | 0.17 | -0.80 | 4.68 | 9.75 | 6.08 | 17.18 |
| | (1.66) | (1.16) | (0.37) | (-1.57) | (1.06) | (1.60) | (0.44) | (0.77) |
| Controls | NO | YES | NO | YES | NO | YES | NO | YES |
| **Panel B: Local** | Equity | | Bonds | | Credits | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | | |
| 1 day | 0.05 | 0.05** | 0.00 | -0.00 | 0.04* | 0.05** | | |
| | (1.51) | (2.25) | (0.01) | (-0.23) | (1.90) | (2.22) | | |
| 1 week | 0.31** | 0.30*** | -0.00 | -0.00 | 0.18* | 0.24** | | |
| | (2.25) | (3.27) | (-0.16) | (-0.42) | (1.66) | (2.09) | | |
| 1 month | 2.25*** | 1.66*** | -0.04 | -0.04 | 0.95** | 1.02** | | |
| | (3.11) | (3.78) | (-0.70) | (-0.96) | (2.22) | (2.38) | | |
| 3 months | 3.54*** | 2.26*** | -0.04 | -0.08 | 2.00 | 2.62* | | |
| | (2.85) | (3.75) | (-0.75) | (-1.37) | (1.52) | (1.75) | | |
| 6 months | 5.37*** | 1.90 | 0.04 | -0.12 | 1.13 | 2.57 | | |
| | (4.21) | (1.20) | (0.39) | (-1.36) | (0.65) | (1.04) | | |
| Controls | NO | YES | NO | YES | NO | YES | | |

Table 8: **Inflation surprises and asset returns: different forecasting horizons.** The table summarizes the results of predictive panel regressions for different forecasting horizons. Regression specifications follow table 4. In Panel A we use global inflation surprises, while in Panel B we use local inflation surprises. The sample period runs from 31-03-1997 until 31-12-2019 and consists of non-overlapping observations. For weekly, monthly, quarterly, or bi-annual forecast horizons we sample at the end of each friday, month, quarter, or December/June cycle, respectively. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| **Panel A: Global** | Equity | | Bonds | | Credits | | Commodities | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 1 day | 0.04 | 0.10** | -0.00** | -0.00** | 0.00 | 0.03* | 0.01 | 0.05 |
| | (1.17) | (1.98) | (-2.23) | (-2.39) | (0.62) | (1.72) | (0.21) | (0.87) |
| 1 week | 0.20 | 0.48 | -0.01* | -0.02** | 0.03 | 0.17 | 0.14 | 0.36 |
| | (0.95) | (1.74) | (-1.77) | (-2.00) | (0.49) | (1.53) | (0.54) | (1.34) |
| 1 month | 1.53** | 2.49*** | -0.05 | -0.08** | 0.42 | 0.99* | 1.31 | 2.25** |
| | (2.35) | (3.07) | (-1.60) | (-2.06) | (1.55) | (1.95) | (1.32) | (2.46) |
| 3 months | 2.29 | 4.91** | -0.17* | -0.34*** | 0.87 | 1.98* | 0.63 | 1.27 |
| | (0.99) | (1.99) | (-1.66) | (-2.79) | (1.01) | (1.76) | (0.23) | (1.47) |
| 6 months | -12.38*** | -11.53** | 0.42** | 0.22 | -2.90* | -0.55 | -9.71* | -5.16 |
| | (-2.70) | (-2.15) | (2.17) | (0.93) | (-1.82) | (-0.32) | (-1.79) | (-1.05) |
| Controls | NO | YES | NO | YES | NO | YES | NO | YES |
| **Panel B: Local** | Equity | | Bonds | | Credits | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | | |
| 1 day | 0.02 | 0.03** | -0.00** | -0.00** | -0.00 | 0.01 | | |
| | (1.38) | (2.02) | (-2.34) | (-2.47) | (-0.28) | (0.83) | | |
| 1 week | 0.10 | 0.17** | -0.01** | -0.01** | -0.03 | 0.01 | | |
| | (1.45) | (2.29) | (-1.98) | (-2.22) | (-0.78) | (0.38) | | |
| 1 month | 0.68*** | 0.91*** | -0.03** | -0.03*** | 0.01 | 0.19 | | |
| | (2.94) | (3.53) | (-2.28) | (-2.63) | (0.05) | (0.95) | | |
| 3 months | 1.19 | 1.85 | -0.09* | -0.15*** | 0.52 | 0.89* | | |
| | (0.92) | (1.59) | (-1.92) | (-2.85) | (1.15) | (1.75) | | |
| 6 months | -4.65** | -2.66 | 0.17 | 0.02 | -1.48* | 0.66 | | |
| | (-2.36) | (-1.57) | (1.49) | (0.37) | (-1.85) | (0.95) | | |
| Controls | NO | YES | NO | YES | NO | YES | | |

Table 9: **Revision versus announcement growth surprises and asset returns.** The table summarizes the impact of revision surprises versus instantaneous announcement growth surprises on the predictive power for asset returns. Regression specifications follow table 4. In Panel A we use instantaneous global growth surprises ('Instantaneous surprise'; hence excluding revision surprises), while in Panel B we use global growth revision surprises ('Revision surprise'; hence excluding instantaneous surprises). The sample period runs from 31-03-1997 until 31-12-2019. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| **Panel A:** | Equity | | Bonds | | Credits | | Commodities | |
|---|---|---|---|---|---|---|---|---|
| **Instantaneous surprise** | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $S_g$ | 14.31*** | 11.36** | −0.27 | −0.42** | 4.31** | 5.01** | 12.60 | 10.70 |
| | (2.96) | (2.44) | (−1.34) | (−2.13) | (2.01) | (2.18) | (1.49) | (1.33) |
| $S_{l-g}$ | 0.82 | 0.79 | 0.07 | 0.07 | 0.78 | 1.18* | | |
| | (0.41) | (0.48) | (0.61) | (0.63) | (1.26) | (1.82) | | |
| Constant | 0.55** | 1.99 | 0.03*** | −0.12 | 0.30** | 0.23 | −0.01 | −1.65 |
| | (2.55) | (1.37) | (3.34) | (−1.48) | (2.15) | (0.44) | (−0.04) | (−0.75) |
| Controls | NO | YES | NO | YES | NO | YES | NO | YES |
| Obs. | 1,056 | 1,044 | 1,056 | 1,044 | 824 | 780 | 274 | 268 |
| $R^2_{adj}$ | 3.1% | 3.8% | 0.6% | 2.9% | 2.4% | 3.3% | 2.9% | 1.6% |
| **Panel B:** | Equity | | Bonds | | Credits | | Commodities | |
| **Revision surprise** | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $S_g$ | 7.45*** | 5.70** | -0.11 | -0.21** | 2.28* | 2.62* | 8.89*** | 9.04*** |
| | (3.15) | (2.40) | (-1.12) | (-2.04) | (1.72) | (1.86) | (3.31) | (3.77) |
| $S_{l-g}$ | 0.68 | 0.72*** | −0.03 | −0.03 | 0.75** | 0.74* | | |
| | (1.10) | (4.40) | (−0.36) | (−0.48) | (2.01) | (1.91) | | |
| Constant | 0.51** | 2.06 | 0.03*** | −0.12 | 0.24* | −0.03 | 0.02 | −1.50 |
| | (2.36) | (1.35) | (3.32) | (−1.48) | (1.91) | (−0.07) | (0.06) | (−0.77) |
| Controls | NO | YES | NO | YES | NO | YES | NO | YES |
| Obs. | 1,056 | 1,044 | 1,056 | 1,044 | 824 | 780 | 274 | 268 |
| $R^2_{adj}$ | 2.5% | 3.3% | 0.2% | 2.5% | 2.2% | 2.9% | 4.6% | 3.5% |

Table 10: **Revision versus announcement inflation surprises and asset returns.** The table summarizes the impact of revision surprises versus instantaneous announcement inflation surprises on the predictive power for asset returns. Regression specifications follow table 5. In Panel A we use instantaneous global growth surprises ('Instantaneous surprise'; hence excluding revision surprises), while in Panel B we use global growth revision surprises ('Revision surprise'; hence excluding instantaneous surprises). The sample period runs from 31-03-1997 until 31-12-2019. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| **Panel A:** | Equity | | Bonds | | Credits | | Commodities | |
|---|---|---|---|---|---|---|---|---|
| **Instantaneous surprise** | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $S_g$ | -1.65 | 0.55 | 0.06** | 0.02 | -0.44 | -0.36 | 1.95 | 1.57 |
| | (-0.68) | (0.19) | (2.00) | (0.17) | (-0.81) | (-0.53) | (0.58) | (0.55) |
| $S_{l-g}$ | -0.18*** | -0.32*** | -0.03 | -0.03 | -0.55 | -0.51 | | |
| | (-4.47) | (-4.09) | (-0.98) | (-1.34) | (-1.27) | (-1.31) | | |
| Constant | 0.43* | 1.59 | 0.03*** | -0.14* | 0.25** | 0.24 | -0.16 | -2.14 |
| | (1.85) | (0.91) | (6.89) | (-1.76) | (2.03) | (0.41) | (-0.51) | (-0.93) |
| Controls | NO | YES | NO | YES | NO | YES | NO | YES |
| Obs. | 1,032 | 982 | 1,032 | 982 | 824 | 780 | 272 | 268 |
| $R^2_{adj}$ | 0.1% | 4.2% | 0.1% | 5.1% | 0.0% | 0.6% | 0.2% | 2.4% |
| **Panel B:** | Equity | | Bonds | | Credits | | Commodities | |
| **Revision surprise** | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $S_g$ | 2.22** | 1.89** | -0.08* | -0.08** | 0.56* | 0.64 | 1.65 | 1.35 |
| | (2.44) | (2.44) | (-1.86) | (-1.97) | (1.65) | (1.61) | (1.44) | (1.44) |
| $S_{l-g}$ | 0.34 | 0.32 | -0.01 | -0.01 | -0.32 | -0.29 | | |
| | (1.49) | (1.18) | (-1.12) | (-1.42) | (-1.35) | (-1.14) | | |
| Constant | 0.45* | 1.66 | 0.03*** | -0.11 | 0.28** | 0.06 | -0.13 | -2.34 |
| | (1.69) | (0.96) | (3.06) | (-1.05) | (2.04) | (0.10) | (-0.43) | (-1.05) |
| Controls | NO | YES | NO | YES | NO | YES | NO | YES |
| Obs. | 1,032 | 1,021 | 1,032 | 1,021 | 824 | 780 | 272 | 268 |
| $R^2_{adj}$ | 1.8% | 4.0% | 1.3% | 2.7% | 1.5% | 2.3% | 1.2% | 3.0% |

Table 11: **Alternative surprise nowcasts: predictive regression results.** The table summarizes the results of alternative methods to combine all individual macroeconomic growth or inflation surprises into a nowcast future on future asset returns. Regression specifications follow table 4 and table 5. 'PCA' columns present the results of the PCA-based method used in previous tables. Alternative weightings schemes include an equal-weighting of each macroeconomic series (EW), an attention-based weighting of macroeconomic series (ATT), and weighting via the three-pass regression filter (3-PRF) of Kelly and Pruitt (2015). Reported are the univariate coefficient estimates, its corresponding t-values, and the $R^2$. The sample period runs from 31-03-1997 until 31-12-2019. The t-values, shown between parenthesis, are adjusted for time-asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| | Growth Surprise | | | | Inflation Surprise | | | |
|---|---|---|---|---|---|---|---|---|
| | PCA | EW | ATT | 3PRF | PCA | EW | ATT | 3PRF |
| Equity | 6.37*** | 4.20*** | 3.02*** | 0.13*** | 1.53** | 1.42*** | 1.60*** | 0.03 |
| | (3.42) | (4.34) | (3.44) | (2.49) | (2.35) | (2.34) | (2.62) | (1.42) |
| Bond | -0.10 | -0.06 | -0.02 | 0.07 | -0.05 | -0.05* | -0.04 | 0.00 |
| | (-1.27) | (-1.24) | (-0.46) | (1.35) | (-1.60) | (-1.75) | (-1.43) | (-0.00) |
| Credits | 1.96* | 0.67* | 0.42* | 0.07 | 0.40 | -0.00 | 1.13 | 0.02 |
| | (1.93) | (1.70) | (1.95) | (1.20) | (1.55) | (0.83) | (1.35) | (0.68) |
| Commodities | 6.94*** | 6.40*** | 3.90*** | 0.16 | 1.31 | 1.30* | 1.01 | 0.00 |
| | (2.78) | (4.32) | (3.71) | (1.46) | (1.32) | (1.70) | (1.45) | (-0.01) |

Table 12: **Alternative surprise nowcasts: strategy results.** This table summarizes the results of alternative methods to combine all individual macroeconomic surprises into a nowcast on investment strategy result. Strategy specifications follow table 6. 'PCA' columns present the results of the PCA-based method used in previous tables. Alternative weightings schemes include equal-weighting each macroeconomic series (EW), an attention-based weighting scheme of macroeconomic series (ATT), and weighting via the three-pass regression filter (3-PRF) of Kelly and Pruitt (2015). We report the Sharpe ratio (Panel A), and the annualized $\hat{\alpha}$ in % (Panel B) of each investment strategy relative to the corresponding global asset class returns from regressing $R_{s,t} = \alpha + \beta R_{m,t}$. The sample runs from 31-03-1997 until 31-12-2019 and consists of monthly non-overlapping observations. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| **Panel A: Sharpe** | Growth Surprise | | | | Inflation Surprise | | | |
|---|---|---|---|---|---|---|---|---|
| | PCA | EW | ATT | 3PRF | PCA | EW | ATT | 3PRF |
| Equity | 0.50** | 0.62*** | 0.65*** | 0.45** | 0.39* | 0.35* | 0.47** | 0.32 |
| | (2.40) | (2.95) | (3.01) | (2.03) | (1.86) | (1.67) | (2.18) | (1.45) |
| Bonds | -0.28 | -0.33 | -0.10 | 0.28 | -0.31 | -0.36* | -0.27 | -0.05 |
| | (-1.33) | (-1.59) | (-0.46) | (1.26) | (-1.46) | (-1.72) | (-1.23) | (-0.21) |
| Credits | 0.44* | 0.48** | 0.45* | 0.16 | 0.37 | 0.28 | 0.31 | 0.09 |
| | (1.86) | (2.03) | (1.90) | (0.63) | (1.58) | (1.17) | (1.30) | (0.35) |
| Commodities | 0.64*** | 0.94*** | 0.86*** | 0.43** | 0.32 | 0.40* | 0.32 | -0.00 |
| | (3.03) | (4.49) | (3.99) | (1.97) | (1.51) | (1.90) | (1.49) | (-0.01) |
| **Panel B: Alpha** | Growth Surprise | | | | Inflation Surprise | | | |
| | PCA | EW | ATT | 3PRF | PCA | EW | ATT | 3PRF |
| Equity | 2.10** | 3.84*** | 4.48*** | 0.55** | 2.48 | 2.37 | 3.46* | 0.56 |
| | (2.35) | (3.42) | (3.69) | (2.11) | (1.56) | (1.48) | (1.88) | (1.42) |
| Bonds | -0.00 | -0.02 | 0.00 | 0.00 | -0.06 | -0.07 | -0.07 | -0.00 |
| | (-0.24) | (-0.82) | (0.06) | (0.93) | (-1.15) | (-1.59 ) | (-1.29) | (-0.04) |
| Credits | 0.56 | 0.87** | 0.91** | 0.08** | 0.69 | 0.39 | 0.46 | 0.06 |
| | (1.37) | (2.07) | (2.15) | (2.48) | (1.39) | (0.90) | (1.00) | (0.92) |
| Commodities | 1.85** | 4.04*** | 4.12*** | 0.41 | 2.00 | 2.37* | 2.18 | 0.00 |
| | (2.37) | (3.64) | (3.33) | (1.22) | (1.21) | (1.74) | (1.38) | (0.00) |

Table 13: **Return predictability and economic surprise momentum: global surprises.** The table summarizes the results of predictive panel regressions split in subsamples based on the sign of current and future global surprises. Regression specifications follow table 4 for growth and table 5 for inflation. "Equal Sign" ("Different Sign") denotes the sample for which the current global surprise index has (not) the same sign as next month (Panel A and B). "Equal Own Sign" ("Different Own Sign") denotes the sample for which the weighted individual surprise has (not) the same sign as next month (Panel C and D). "Equal Cross Sign" ("Different Cross Sign") denotes the sample for which the weighted cross terms have (not) the same sign as next month (Panel C and D). Shown are the predictive regression estimates, its corresponding t-values, the number of observations (Obs.), and the adjusted $R^2$. The sample period runs from 31-03-1997 until 31-12-2019. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| **Panel A: Growth** | Equity | | Bonds | | Credits | | Commodity | |
|---|---|---|---|---|---|---|---|---|
| | $S_g$ | $R^2$ | $S_g$ | $R^2$ | $S_g$ | $R^2$ | $S_g$ | $R^2$ |
| Equal Sign | 7.10*** | 4.8% | -0.16* | 1.6% | 2.76** | 6.8% | 6.70** | 5.0% |
| | (3.30) | | (-1.71) | | (2.07) | | (2.29) | |
| Different Sign | 2.49 | 0.1% | 0.17 | 0.6% | -2.10 | 1.3% | 7.83** | 4.6% |
| | (0.88) | | (1.23) | | (-1.50) | | (2.08) | |
| Full Sample | 6.37*** | 3.6% | -0.10 | 0.4% | 1.96* | 2.7% | 6.94*** | 5.3% |
| | (3.42) | | (-1.27) | | (1.93) | | (2.78) | |
| **Panel B: Inflation** | Equity | | Bonds | | Credits | | Commodity | |
| | $S_g$ | $R^2$ | $S_g$ | $R^2$ | $S_g$ | $R^2$ | $S_g$ | $R^2$ |
| Equal Sign | 1.67* | 1.2% | -0.07 | 0.9% | 0.44 | 0.7% | 0.97 | -0.2% |
| | (1.68) | | (-1.21) | | (1.28) | | (0.72) | |
| Different Sign | 1.40* | 0.8% | -0.04 | 0.4% | 0.44 | 0.7% | 1.86* | 1.4% |
| | (1.67) | | (-1.32) | | (1.36) | | (1.89) | |
| Full Sample | 1.53** | 1.1% | -0.05 | 0.8% | 0.42 | 0.7% | 1.31 | 0.7% |
| | (2.35) | | (-1.60) | | (1.55) | | (1.32) | |
| **Panel C: Growth** | Equity | | Bonds | | Credits | | Commodity | |
| | $S_g$ | $R^2$ | $S_g$ | $R^2$ | $S_g$ | $R^2$ | $S_g$ | $R^2$ |
| Equal Own Sign | 6.78*** | 4.4% | -0.12 | 0.6% | 2.26* | 3.9% | 6.81** | 5.0% |
| | (3.46) | | (-1.32) | | (1.96) | | (2.43) | |
| Different Own Sign | 3.75 | 0.4% | -0.02 | -0.4% | 0.47 | -0.3% | 8.05** | 5.2% |
| | (0.98) | | (-0.15) | | (0.48) | | (2.16) | |
| Equal Cross Sign | 6.40*** | 4.4% | -0.15 | 1.2% | 2.18* | 4.4% | 9.10*** | 10.0% |
| | (3.05) | | (-1.53) | | (1.89) | | (3.06) | |
| Different Cross Sign | 6.64** | 2.2% | 0.04 | -0.2% | 1.10 | 0.1% | 0.35 | -1.0% |
| | (2.02) | | (0.31) | | (0.73) | | (0.09) | |
| **Panel D: Inflation** | Equity | | Bonds | | Credits | | Commodity | |
| | $S_g$ | $R^2$ | $S_g$ | $R^2$ | $S_g$ | $R^2$ | $S_g$ | $R^2$ |
| Equal Own Sign | 1.65 | 1.2% | 0.01 | -0.2% | 0.29 | 0.2% | 2.10 | 1.7% |
| | (1.43) | | (0.26) | | (0.77) | | (1.16) | |
| Different Own Sign | 1.39* | 0.8% | -0.10** | 2.9% | 0.49 | 1.0% | 0.66 | -0.3% |
| | (1.72) | | (-2.31) | | (1.41) | | (0.58) | |
| Equal Cross Sign | 1.48 | 0.7% | -0.01 | -0.2% | 0.06 | -0.2% | 2.40* | 2.2% |
| | (1.35) | | (-0.19) | | (0.18) | | (1.80) | |
| Different Cross Sign | 1.39* | 1.0% | -0.09** | 2.3% | 0.58 | 1.6% | 0.23 | -0.7% |
| | (1.78) | | (-1.98) | | (1.60) | | (0.28) | |

Table 14: **Growth surprises and asset volatility.** This table summarizes the impact of macroeconomic growth surprises on asset volatility. We regress one-month ahead realized volatility on end-of-month growth surprises for different asset classes. For the remainder, regression specifications and definitions follow table 4. Shown are the predictive regression estimates, its corresponding t-values (in parenthesis), the number of observations (Obs.), and the adjusted $R^2$. The sample period runs from 31-03-1997 until 31-12-2019. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| | Equity | | | | Bonds | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $S_l$ | −0.35 | | | | −0.00 | | | |
| | (−1.42) | | | | (−1.68) | | | |
| $S_g$ | | −1.23*** | −1.27*** | −0.73** | | −0.00** | −0.00** | −0.00** |
| | | (−3.16) | (−3.05) | (−2.21) | | (−2.48) | (−2.47) | (−2.29) |
| $S_{l-g}$ | | | 0.03 | 0.02 | | | −0.00 | −0.00 |
| | | | (0.09) | (0.27) | | | (−1.17) | (−1.19) |
| C | 0.37*** | 0.35*** | 0.34*** | 0.03 | 0.00*** | 0.00*** | 0.00*** | −0.00*** |
| | (5.68) | (6.50) | (6.50) | (0.15) | (4.27) | (4.51) | (4.25) | (−2.72) |
| Controls | NO | NO | NO | YES | NO | NO | NO | YES |
| Obs. | 1,056 | 1,096 | 1,056 | 1,044 | 1,056 | 1,096 | 1,056 | 1,044 |
| $R^2_{adj}$ | 2.2% | 8.3% | 8.4% | 18.5% | 3.7% | 3.8% | 4.9% | 37.4% |
| | Credits | | | | Commodities | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | | |
| $S_l$ | −0.03 | | | | | | | |
| | (−1.55) | | | | | | | |
| $S_g$ | | −0.04* | −0.04** | 0.00 | −0.29 | −0.11 | | |
| | | (−1.90) | (−2.12) | (0.20) | (−0.98) | (−0.80) | | |
| $S_{l-g}$ | | | −0.02 | −0.02 | | | | |
| | | | (−1.11) | (−1.20) | | | | |
| C | 0.03** | 0.03** | 0.03** | −0.02** | 0.19*** | −0.13** | | |
| | (2.07) | (2.05) | (2.06) | (−2.82) | (9.83) | (−1.42) | | |
| Controls | NO | NO | NO | YES | NO | YES | | |
| Obs. | 824 | 824 | 824 | 780 | 274 | 268 | | |
| $R^2_{adj}$ | 2.5% | 1.8% | 2.7% | 20.4% | 4.9% | 42.9% | | |

Table 15: **Inflation surprises and asset volatility.** This table summarizes the impact of macroeconomic inflation surprises on asset volatility. We regress one-month ahead realized volatility on end-of-month inflation surprises for different asset classes. For the remainder, regression specifications and definitions follow table 5. Shown are the predictive regression estimates, its corresponding t-values (in parenthesis), the number of observations (Obs.), and the adjusted $R^2$. The sample period runs from 31-03-1997 until 31-12-2019. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| | Equity | | | | Bonds | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $S_l$ | $-0.07$ | | | | $-0.00$ | | | |
| | $(-1.22)$ | | | | $(-0.21)$ | | | |
| $S_g$ | | $-0.22^{**}$ | $-0.22^{**}$ | $-0.19$ | | $-0.00^{*}$ | $-0.00^{*}$ | $-0.00$ |
| | | $(-2.60)$ | $(-2.55)$ | $(-1.36)$ | | $(-1.66)$ | $(-1.93)$ | $(-0.92)$ |
| $S_{l-g}$ | | | $0.02$ | $0.03$ | | | $0.00^{***}$ | $0.00$ |
| | | | $(0.32)$ | $(0.31)$ | | | $(3.00)$ | $(1.64)$ |
| C | $0.37^{***}$ | $0.38^{***}$ | $0.37^{***}$ | $-0.08$ | $0.00^{***}$ | $0.00^{***}$ | $0.00^{***}$ | $-0.00^{**}$ |
| | $(6.74)$ | $(6.52)$ | $(6.61)$ | $(-0.44)$ | $(4.21)$ | $(4.38)$ | $(4.24)$ | $(-2.13)$ |
| Controls | NO | NO | NO | YES | NO | NO | NO | YES |
| Obs. | 1,032 | 1,088 | 1,032 | 982 | 1,032 | 1,088 | 1,032 | 982 |
| $R^2_{adj}$ | 0.3% | 1.4% | 1.3% | 18.6% | −0.1% | 0.5% | 0.9% | 38.5% |
| | Credits | | | | Commodity | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | | |
| $S_l$ | $-0.01$ | | | | | | | |
| | $(-1.31)$ | | | | | | | |
| $S_g$ | | $-0.01$ | $-0.01$ | $0.00$ | $-0.09$ | $-0.08^{*}$ | | |
| | | $(-1.54)$ | $(-1.51)$ | $(0.11)$ | $(-1.50)$ | $(-1.93)$ | | |
| $S_{l-g}$ | | | $-0.00$ | $0.00$ | | | | |
| | | | $(-0.49)$ | $(0.02)$ | | | | |
| C | $0.03^{**}$ | $0.03^{**}$ | $0.03^{**}$ | $-0.04^{***}$ | $0.19^{***}$ | $-0.12^{**}$ | | |
| | $(2.05)$ | $(2.05)$ | $(2.05)$ | $(-2.93)$ | $(8.14)$ | $(-2.12)$ | | |
| Controls | NO | NO | NO | YES | NO | YES | | |
| Obs. | 824 | 824 | 824 | 780 | 272 | 255 | | |
| $R^2_{adj}$ | 0.3% | 0.6% | 0.5% | 20.0% | 2.5% | 48.2% | | |

Table 16: **Negative expected return tests: macroeconomic growth surprises.** This table summarizes the results of multiple tests regarding negative expected returns and global growth surprises. Panel A reports the number (#) and fraction (%) of expected return forecasts for which the 90% confidence interval lies below zero. Panel B reports the estimate of the non-linearity regression using global growth surprises. $D$ takes value one if expected excess returns are negative, and zero otherwise. The last row report the Wald test on being statistically significantly different from zero ($\chi^2(1)$). The p-values are computed with double-clustered (by time and asset) standard errors. Panel C reports the minimum expected return and its p-value based (within parenthesis) on the minimum expected return test of Eleswarapu and Thompson (2007) using 10.000 bootstraps. The sample period runs from 31-03-1997 until 31-12-2019. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| **Panel A** | Equity | Bonds | Credits | Commodities |
|---|---|---|---|---|
| $E(r) < 0$ | 224 of 1,096 | 0 of 1,096 | 40 of 824 | 99 of 274 |
| | (20.4%) | (0.0%) | (4.9%) | (36.1%) |
| **Panel B** | | | | |
| $\alpha$ | 0.72 | 0.03 | 0.34 | 0.26 |
| | (2.56) | (3.44) | (2.30) | (0.66) |
| $G_{sur}$ | 4.57* | -0.12 | 0.96 | 4.84 |
| | (1.84) | (1.47) | (1.10) | (1.52) |
| $D \times G_{sur}$ | 3.14 | 0.20 | 2.14 | 3.53 |
| | (0.74) | (0.92) | (1.58) | (0.55) |
| $G_{sur} + D \times G_{sur}$ | 7.71*** | 0.08** | 3.10*** | 8.37*** |
| | (21.73) | (3.99) | (15.00) | (8.29) |
| **Panel C** | | | | |
| Min. $E(r)(\%)$ | -3.12*** | -0.01 | -0.85** | -3.58*** |
| | (0.00) | (0.65) | (0.01) | (0.00) |

Table 17: **Negative expected return tests: macroeconomic inflation surprises.** This table summarizes the results of multiple tests regarding negative expected returns and global inflation surprises. Panel A reports the number (#) and fraction (%) of expected return forecasts for which the 90% confidence interval lies below zero. Panel B reports the estimate of the non-linearity regression using global inflation surprises. $D$ takes value one if expected excess returns are negative, and zero otherwise. The last row report the Wald test on being statistically significantly different from zero ($\chi^2(1)$). The p-values are computed with double-clustered (by time and asset) standard errors. Panel C reports the minimum expected return and its p-value based (within parenthesis) on the minimum expected return test of Eleswarapu and Thompson (2007) using 10.000 bootstraps. The sample period runs from 31-03-1997 until 31-12-2019. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| **Panel A** | Equity | Bonds | Credits | Commodities |
|---|---|---|---|---|
| $E(r) < 0$ | 72 of 1088 | 0 of 1088 | 0 of 824 | 35 of 272 |
| | (6.6%) | (0.0%) | (0.0%) | (12.9%) |
| **Panel B** | | | | |
| $\alpha$ | 0.44** | 0.03*** | 0.28 | 0.01 |
| | (2.03) | (3.57) | (2.05) | (0.02) |
| $G_{sur}$ | 1.55** | -0.04 | 0.77 | 0.63 |
| | (2.53) | (-0.79) | (1.59) | (0.32) |
| $D \times G_{sur}$ | 2.42* | 0.10 | 0.80 | 1.19 |
| | (1.89) | (1.16) | (1.16) | (0.38) |
| $G_{sur} + D \times G_{sur}$ | 3.98*** | 0.06 | 1.57*** | 1.82 |
| | (10.44) | (1.65) | (11.40) | (1.62) |
| **Panel C** | | | | |
| Min. $E(r)(\%)$ | -1.93*** | -0.01 | -0.54* | -1.80* |
| | (0.01) | (0.27) | (0.08) | (0.06) |

## 2.9 Appendix

Table A.1: **Macroeconomic announcement series: U.S.** This table reports the macroeconomic series included in our sample for the US, the category to which they belong (Cat.), the release frequency of the series (Freq), whether we applied first-differencing (FD) or seasonal-differencing (Seas.), and the total number of surprise observations (Obs.) for each series included in our sample.

| Variable (U.S.) | Cat. | Freq | FD | Seas. | Obs. | Variable | Cat. | Freq | FD | Seas. | Obs. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Change Nonfarm Payrolls | empl | M | NO | NO | 276 | Building Permits | out | M | YES | NO | 209 |
| Initial Jobless Claims | empl | W | YES | NO | 1,189 | Capacity Utilization | out | M | NO | NO | 276 |
| Unemployment Rate | empl | M | YES | NO | 276 | Cap Goods Orders Nondef ex. Air | out | M | NO | NO | 80 |
| ADP Employment Change | empl | M | YES | NO | 160 | S&P/CS HPI Composite (YoY) | out | M | YES | NO | 152 |
| Change Manuf. Payrolls | empl | M | NO | NO | 252 | NAHB Housing Market Index | out | M | YES | NO | 200 |
| Continuing Claims | empl | W | YES | NO | 868 | Total Vehicle Sales | out | M | YES | NO | 203 |
| Avg. Hourly Earnings (MoM) | empl | M | YES | YES | 118 | Consumer Credit | out | M | YES | NO | 278 |
| Avg. Hourly Earnings (YoY) | empl | M | YES | NO | 118 | Business Inventories | out | M | NO | NO | 271 |
| Avg Weekly Hours | empl | M | YES | NO | 118 | Dom. Vehicle Sales | out | M | YES | NO | 226 |
| GDP (QoQ) | out | Q | NO | NO | 92 | S&P/CS 20-City (MoM) | out | M | YES | NO | 121 |
| Ism Manufacturing PMI | out | M | NO | NO | 278 | Mortgage Application | out | M | NO | NO | 21 |
| Dur. Goods Orders | out | M | NO | NO | 251 | Adj. Retail Sales ex. Autos | out | M | NO | NO | 223 |
| New Home Sales | out | M | YES | NO | 260 | CCI (U.S.) | sent | M | YES | NO | 274 |
| Retail Sales (MoM) | out | M | NO | NO | 223 | Michigan CSI | sent | M | YES | NO | 247 |
| Housing Starts | out | M | YES | NO | 262 | Empire Manuf. Survey | sent | M | NO | NO | 205 |
| Housing Starts (MoM) | out | M | NO | NO | 124 | Chicago PMI | sent | M | NO | NO | 274 |
| Ind. Prod (MoM) | out | M | NO | NO | 278 | Philly Business Outlook | sent | M | NO | NO | 275 |
| Existing Home Sales | out | M | YES | NO | 178 | NFIB Small Business Optimism | sent | M | YES | NO | 118 |
| Factory Orders | out | M | NO | NO | 279 | Richmond Manufacturing Survey | sent | M | NO | NO | 169 |
| Personal Income | out | M | NO | NO | 278 | CPI (MoM) | infl | M | NO | NO | 277 |
| Personal Spending | out | M | NO | NO | 276 | PPI Final Demand (MoM) | infl | M | NO | NO | 71 |
| Trade Balance | out | M | YES | NO | 278 | GDP Price Index | infl | Q | NO | NO | 59 |
| Constr Spending (MoM) | out | M | NO | NO | 198 | Import Price Index (MoM) | infl | M | NO | NO | 257 |
| Pending Home Sales (MoM) | out | M | NO | NO | 176 | CPI ex. Food&Energy (MoM) | infl | M | NO | NO | 274 |
| Pending Home Sales (YoY) | out | M | YES | NO | 87 | CPI ex. Food&Energy (YoY) | infl | M | YES | NO | 194 |
| Monthly Budget Statement | out | M | YES | NO | 278 | Employment Cost Index | infl | Q | YES | NO | 84 |
| ISM Nonmanuf. Comp. | out | M | YES | NO | 143 | ISM Prices Paid | infl | M | NO | NO | 234 |
| Durables ex. Transport | out | M | NO | NO | 186 | PPI ex. Food&Energy (MoM) | infl | M | NO | NO | 71 |
| Current Acc Balance | out | Q | YES | NO | 87 | PCE core (MoM) | infl | M | YES | NO | 175 |
| Personal Consumption | out | Q | NO | NO | 68 | NonFarm Productivity | infl | Q | NO | NO | 88 |
| FHFA House Price Index (MoM) | out | M | YES | NO | 141 | PCE Core (YoY) | infl | M | YES | NO | 183 |

Table A.2: **Macroeconomic announcement series: U.K..** This table reports the macroeconomic series included in our sample for the UK, the category to which they belong (Cat.), the release frequency of the series (Freq), whether we applied first-differencing (FD) or seasonal-differencing (Seas.), and the total number of surprise observations (Obs.) for each series included in our sample.

| Variable (U.K.) | Cat. | Freq | FD | Seas. | Obs. | Variable | Cat. | Freq | FD | Seas. | Obs. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Jobless Claims Change | empl | M | NO | NO | 240 | Net Consumer Credit | out | M | YES | NO | 88 |
| ILO Unemployment Rate | empl | M | NO | NO | 212 | Business Investments (QoQ) | out | Q | NO | NO | 47 |
| Weekly Earnings Ex Bon3M (YoY) | empl | M | YES | NO | 120 | Net Lending on Dwellings | out | M | YES | NO | 211 |
| Earnings Growth 3M | empl | M | YES | NO | 120 | GFK Consumer Conf. | sent | M | YES | NO | 204 |
| Claimant Rate | empl | M | NO | NO | 234 | CBI Total Orders Book Balance | sent | M | NO | NO | 113 |
| GDP (QoQ) | out | Q | NO | NO | 85 | CBI Reported Sales | sent | M | NO | NO | 112 |
| Nationwide House Price (MoM) | out | M | NO | NO | 187 | CBI Selling Prices | sent | M | NO | NO | 64 |
| Ind. Prod (Mom) | out | M | NO | NO | 276 | CPI (YoY) | infl | M | YES | NO | 199 |
| House Price (MoM) (YoY) | out | M | NO | NO | 187 | CPI (MoM) | infl | M | NO | YES | 192 |
| PPI Output (MoM) | out | M | NO | NO | 271 | RPI (MoM) | infl | M | NO | YES | 275 |
| Mortgage Approvals | out | M | YES | NO | 181 | Money Supply M4 (MoM) | infl | M | NO | NO | 171 |
| Retail Sales Ex. Auto (MoM) | out | M | NO | NO | 274 | PPI Plusfuel | infl | M | NO | NO | 172 |
| Trade Balance | out | M | YES | NO | 160 | RPI Xmortg | infl | M | NO | NO | 276 |
| Index Total Service 3M | out | M | NO | NO | 150 | | | | | | |

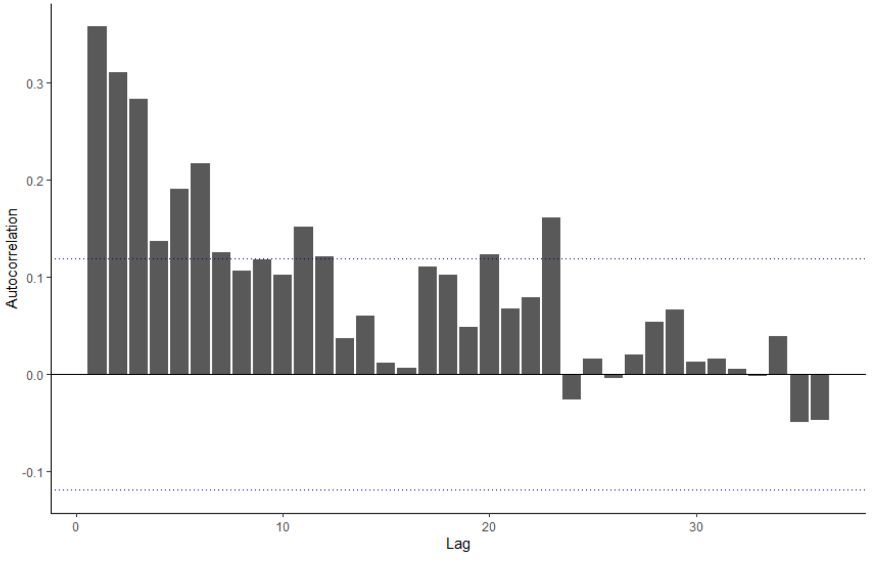
Table A.3: **Macroeconomic announcement series: Japan.** This table reports the macroeconomic series included in our sample for Japan, the category to which they belong (Cat.), the release frequency of the series (Freq), whether we applied first-differencing (FD) or seasonal-differencing (Seas.), and the total number of surprise observations (Obs.) for each series included in our sample.

| Variable (Japan) | Cat. | Freq | FD | Seas. | Obs. | Variable | Cat. | Freq | FD | Seas. | Obs. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Job To Applicant Ratio | empl | M | YES | NO | 239 | Exports (YoY) | out | M | NO | NO | 130 |
| Labor Cash Earnings (YoY) | empl | M | NO | NO | 152 | Real GDP (QoQ) | out | Q | NO | NO | 61 |
| Unemployment Rate | empl | M | YES | NO | 239 | Tankan Large Mfg Index | sent | Q | NO | NO | 85 |
| Industrial Production (MoM) | out | M | NO | NO | 221 | Eco Watchers Survey Current | sent | M | YES | NO | 104 |
| Gdp (QoQ) | out | Q | NO | NO | 61 | Small Business Confidence | sent | M | NO | NO | 64 |
| Tertiary Industry Index (MoM) | out | M | NO | YES | 238 | Eco Watchers Survey Outlook | sent | M | NO | NO | 46 |
| All Industry Activity Index (MoM) | out | M | NO | YES | 199 | Tokyo CPI Ex Fresh Food (YoY) | infl | M | NO | NO | 218 |
| Capital Spending (YoY) | out | Q | YES | NO | 58 | Domestic Cgpi (YoY) | infl | M | NO | NO | 195 |
| Ind. Prod. (YoY) | out | M | NO | NO | 189 | GDP Deflator (YoY) | infl | Q | YES | NO | 60 |
| Machine Orders (YoY) | out | M | NO | NO | 197 | CPI (YoY) | infl | M | NO | NO | 220 |
| Housing Starts (YoY) | out | M | NO | NO | 235 | CPI Ex Fresh Food (YoY) | infl | M | NO | NO | 219 |
| Trade Balance Bop Basis | out | M | YES | YES | 201 | Tokyo CPI (YoY) | infl | M | YES | NO | 220 |
| Retail Trade (YoY) | out | M | YES | NO | 199 | CPI Ex Food Energy (YoY) | infl | M | YES | NO | 117 |
| Overall Household Spending (YoY) | out | M | NO | NO | 164 | Tokyo CPI Ex Food Energy (YoY) | infl | M | YES | NO | 112 |
| Bank Lending Incl Trusts (YoY) | out | M | YES | NO | 92 | Corp Serv | infl | M | YES | NO | 208 |

Table A.4: **Macroeconomic announcement series: Eurozone.** This table reports the macroeconomic series included in our sample for the Eurozone countries, the category to which they belong (Cat.), the release frequency of the series (Freq), whether we applied first-differencing (FD) or seasonal-differencing (Seas.), and the total number of surprise observations (Obs.) for each series included in our sample.

| Variable (Europe) | Cat. | Freq | FD | Seas. | Obs. | Variable | Cat. | Freq | FD | Seas. | Obs. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Unemployment Change (GE) | empl | M | NO | NO | 254 | Retail Sales (MoM,GE) | out | M | NO | NO | 255 |
| Unemployment Rate (EC) | empl | M | YES | NO | 236 | Industrial Orders (MoM,IT) | out | M | NO | YES | 117 |
| Job Seek Chng (FR) | empl | M | YES | NO | 88 | GDP (QoQ,SP) | out | Q | YES | NO | 69 |
| Pay Nonfarm (FR) | empl | Q | NO | NO | 57 | Adj. Ind. Prod. (YoY,SP) | out | M | NO | NO | 72 |
| Unmp All (FR) | empl | Q | YES | NO | 45 | Zew Survey Exp. (GE) | sent | M | NO | NO | 213 |
| Unmp Mainl (FR) | empl | Q | YES | NO | 43 | Business Conf. (IT) | sent | M | YES | NO | 232 |
| Wage Hmom (IT) | empl | M | YES | NO | 61 | IFO Business Climate (GE) | sent | M | YES | NO | 179 |
| Unempl. Rate (IT) | empl | M | YES | NO | 119 | Business Confidence (BE) | sent | M | NO | NO | 207 |
| Unempl. Lvl. (MoM,SP) | empl | M | YES | YES | 167 | Consumer Confidence (EC) | sent | M | NO | NO | 201 |
| Unempl. Rate (SP) | empl | Q | YES | NO | 71 | Consumer Confidence (FR) | sent | M | YES | NO | 103 |
| Labor Cost (EU) | empl | Q | YES | NO | 15 | Manufacturing Confidence (FR) | sent | M | NO | NO | 218 |
| GDP (QoQ,EC) | out | Q | NO | NO | 56 | GFK Consumer Confidence (GE) | sent | M | YES | NO | 158 |
| Ind. Prod. (MoM,GE) | out | M | NO | NO | 273 | Economic Confidence (EC) | sent | M | NO | NO | 201 |
| Ind. Prod. (MoM,IT) | out | M | NO | NO | 227 | Bank Of France Business Sentiment (FR) | sent | M | NO | NO | 136 |
| Ind. Prod. (MoM,FR) | out | M | NO | NO | 274 | Business Climate Indicator (EC) | sent | M | NO | NO | 206 |
| Ind. Prod. (YoY,FR) | out | M | NO | NO | 272 | Industrial Confidence (EC) | sent | M | NO | NO | 201 |
| Factory Orders (MoM,GE) | out | M | NO | NO | 253 | Production Outlook Indicator (FR) | sent | M | NO | NO | 187 |
| GDP (QoQ,IT) | out | Q | YES | NO | 65 | Own Company Production Outlook (FR) | sent | M | NO | NO | 75 |
| GDP (QoQ,FR) | out | Q | NO | NO | 74 | Sentix Investsor Confidence (EC) | sent | M | YES | NO | 140 |
| GDP (QoQ,GE) | out | Q | NO | NO | 78 | IFO Pan Exp (GE) | sent | M | YES | NO | 179 |
| Retail Sales (MoM,IT) | out | M | YES | NO | 174 | CCI (IT) | sent | M | YES | NO | 220 |
| Retail Sales (MoM,EC) | out | M | NO | NO | 226 | Conf. Index (EU) | sent | M | YES | NO | 177 |
| Retail Sales (YoY,EC) | out | M | YES | NO | 223 | Growth Expectations (EU) | sent | M | NO | NO | 85 |
| Manuf. Prod (MoM,FR) | out | M | YES | YES | 264 | M3 Money Supply (YoY,EC) | infl | M | NO | NO | 220 |
| Trade Balance (GE) | out | M | YES | NO | 216 | CPI (YoY,EC) | infl | M | YES | NO | 214 |
| Industrial Production (YoY,EC) | out | M | NO | NO | 225 | CPI Har (MoM,SP) | infl | M | NO | YES | 186 |
| Trade Balance (EC) | out | M | YES | NO | 160 | PPI(MoM,SP) | infl | M | NO | NO | 97 |
| Consumer Spending (MoM,FR) | out | M | NO | NO | 96 | CPI (MoM,EC) | infl | M | NO | YES | 223 |
| Consumer Spending (YoY,FR) | out | M | NO | NO | 96 | CPI (MoM,GE) | infl | M | NO | YES | 203 |
| GDP Dom. Demand (GE) | out | Q | NO | NO | 53 | CPI (MoM,FR) | infl | M | NO | YES | 273 |
| GDP Expenditures (GE) | out | Q | NO | NO | 61 | PPI (MoM,EC) | infl | M | NO | NO | 224 |
| GDP Gov. Consumption (GE) | out | Q | NO | NO | 63 | CPI (IT) | infl | M | YES | YES | 106 |
| GDP Capital Inv. (GE) | out | Q | NO | NO | 48 | PPI (MoM,FR) | infl | M | NO | NO | 162 |
| GDP Import (GE) | out | Q | NO | NO | 61 | PPI (MoM,GE) | infl | M | NO | NO | 276 |
| GDP Inv. Construction (GE) | out | Q | NO | NO | 59 | HICP (MoM,IT) | infl | M | NO | YES | 205 |
| GDP Priv. Consumption (GE) | out | Q | NO | NO | 65 | CPI (MoM,SP) | infl | M | NO | YES | 209 |

Figure A.5: **Economic surprise momentum using monthly series.** The upper (lower) plot depicts the autocorrelation pattern of the global growth (inflation) surprise nowcast. To construct nowcasts, we only use macroeconomic series that are released on a monthly frequency. Autocorrelations are calculated at the monthly frequency (using 21 business-day sub-sampling) for lags 1 to 36. Dotted lines indicate 5% significance levels. The sample runs from 31-03-1997 until 31-12-2019.



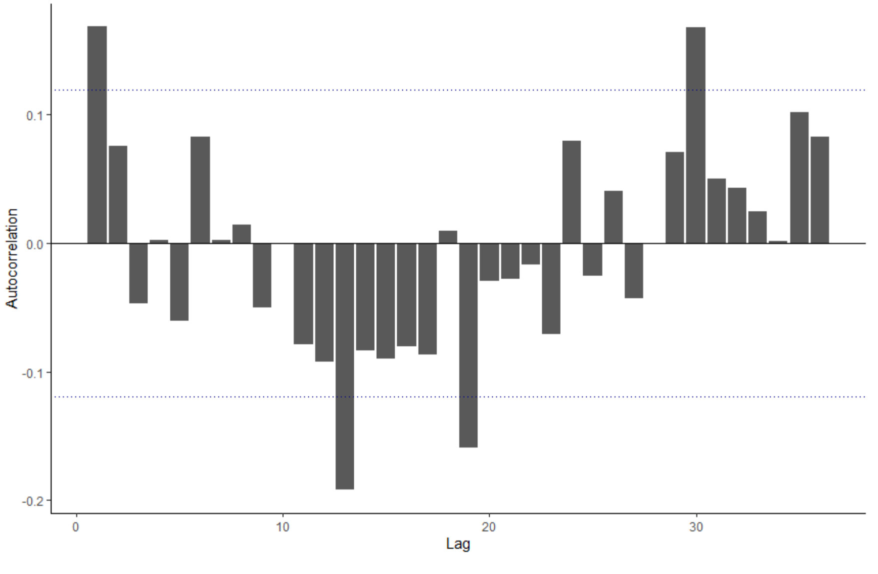
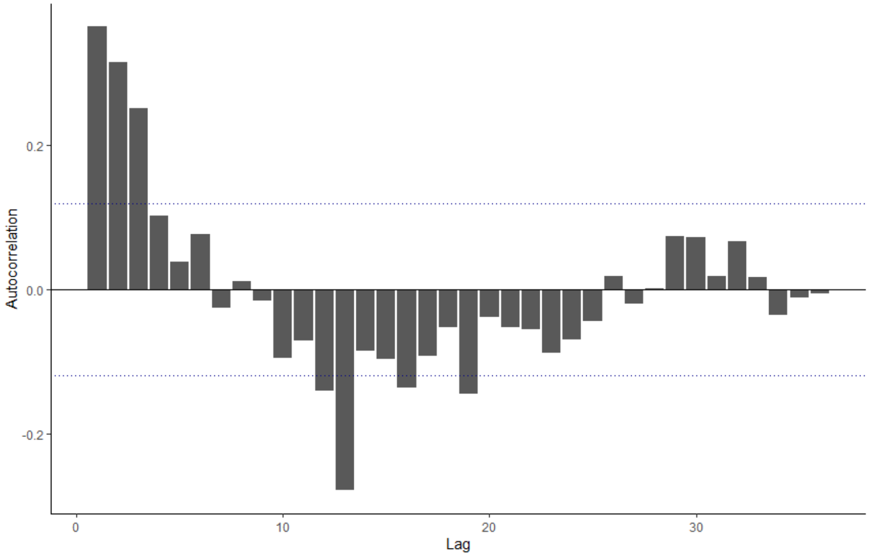



Figure A.6: **Economic surprise momentum using monthly series and excluding stale forecasts.** The upper (lower) plot depicts the autocorrelation pattern of the global growth (inflation) surprise nowcast. To construct nowcasts, we only use macroeconomic series that are released on a monthly frequency. We exclude observations whereby the consensus forecast is stale (no change in the forecast relative to the previous forecast). Autocorrelations are calculated at the monthly frequency (using 21 business-day sub-sampling) for lags 1 to 36. Dotted lines indicate 5% significance levels. The sample runs from 31-03-1997 until 31-12-2019.



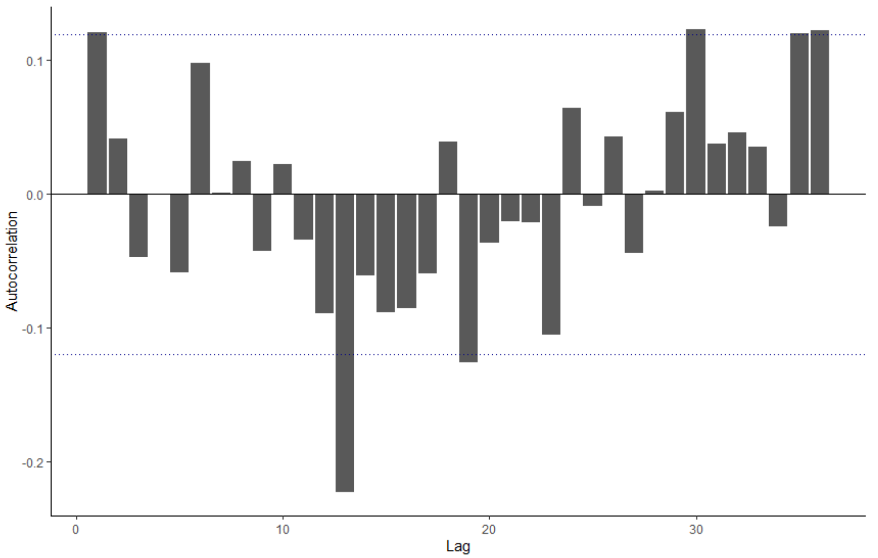



Table A.7: **Macroeconomic surprise strategy: asset level results.** This table summarizes the results of macroeconomic growth and inflation nowcast investment strategies per asset included in our sample. Strategy specifications follow table 6. We report the Sharpe ratio, the annualized $\hat{\alpha}$ in %, and the market exposure ($\hat{\beta}$) of each investment strategy relative to the corresponding global asset class returns from regressing $R_{s,t} = \alpha + \beta R_{m,t}$. The sample runs from 31-03-1997 until 31-12-2019 and consists of monthly non-overlapping observations. The t-values, shown between parenthesis, are Newey-West corrected. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| | Growth Surprise | | | Inflation Surprise | | |
|---|---|---|---|---|---|---|
| Equity | Sharpe | $\hat{\alpha}$ | $\hat{\beta}$ | Sharpe | $\hat{\alpha}$ | $\hat{\beta}$ |
| US | 0.51** | 2.67** | -0.12** | 0.35* | 2.42 | -0.06 |
| | (2.43) | (2.41) | (-2.30) | (1.66) | (1.38) | (-1.19) |
| UK | 0.70*** | 2.16*** | -0.07** | 0.37* | 2.18 | -0.05 |
| | (3.32) | (2.71) | (-1.99) | (1.75) | (1.63) | (-1.03) |
| JP | 0.25 | 1.32 | -0.06* | 0.35 | 2.47 | -0.01 |
| | (1.21) | (1.59) | (-1.87) | (1.64) | (1.56) | (-0.19) |
| EU | 0.47** | 2.17** | -0.09** | 0.37* | 2.89 | -0.05 |
| | (2.24) | (2.11) | (-2.49) | (1.75) | (1.41) | (-1.08) |
| | Growth Surprise | | | Inflation Surprise | | |
| Bonds | Sharpe | $\hat{\alpha}$ | $\hat{\beta}$ | Sharpe | $\hat{\alpha}$ | $\hat{\beta}$ |
| US 10Y | -0.22 | -0.00 | -0.11 | -0.31 | -0.10 | -0.08 |
| | (-1.06) | (-0.08) | (-1.47) | (-1.45) | (-1.49) | (-0.88) |
| UK 10Y | -0.25 | -0.02 | -0.07 | -0.25 | -0.06 | -0.10 |
| | (-1.21) | (-0.59) | (-1.57) | (-1.21) | (-0.91) | (-0.94) |
| JP 10Y | -0.25 | -0.00 | -0.07*** | -0.15 | -0.03 | 0.02 |
| | (-1.18) | (-0.08) | (-2.78) | (-0.72) | (-0.69) | (0.43) |
| EU 10Y | -0.29 | -0.01 | -0.07* | -0.31 | -0.06 | -0.06 |
| | (-1.36) | (-0.26) | (-1.83) | (-1.46) | (-1.15) | (-1.14) |
| | Growth Surprise | | | Inflation Surprise | | |
| Credits | Sharpe | $\hat{\alpha}$ | $\hat{\beta}$ | Sharpe | $\hat{\alpha}$ | $\hat{\beta}$ |
| USIG | 0.46* | 0.35** | -0.13* | 0.02 | 0.05 | -0.03 |
| | (1.89) | (2.11) | (-1.86) | (0.06) | (0.30) | (-0.29) |
| USHY | 0.48** | 1.16 | -0.05 | 0.28 | 1.49* | -0.06 |
| | (2.03) | (1.46) | (-0.68) | (1.21) | (1.79) | (-0.64) |
| EUIG | 0.26 | 0.15 | -0.05 | 0.30 | 0.25 | 0.02 |
| | (1.10) | (1.18) | (-1.32) | (1.27) | (1.12) | (0.36) |
| EUHY | 0.45* | 0.79 | -0.02 | 0.39 | 1.57 | -0.04 |
| | (1.79) | (1.21) | (-0.49) | (1.55) | (1.47) | (-0.61) |

Table A.8: **Growth surprises and asset skewness.** This table summarizes the impact of macroeconomic growth surprises on asset skewness. We regress one-month ahead realized skewness on end-of-month growth surprises for different asset classes. For the remainder, regression specifications and definitions follow table 4. Shown are the predictive regression estimates, its corresponding t-values, the number of observations (Obs.), and the adjusted $R^2$. The sample period runs from 31-03-1997 until 31-12-2019. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| | Equity | | | | Bonds | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $S_l$ | −2.60 | | | | −4.78 | | | |
| | (−0.53) | | | | (−1.35) | | | |
| $S_g$ | | −6.53 | −9.09 | 7.623 | | −10.32 | −11.21 | −17.70 |
| | | (−0.55) | (−0.72) | (0.51) | | (−0.53) | (−0.58) | (−0.97) |
| $S_l$ | | | 0.03 | 3.22 | | | −2.17 | −2.04 |
| | | | (0.02) | (0.75) | | | (−0.77) | (−0.65) |
| C | 0.57 | 0.75 | 0.39 | −18.45* | −5.17*** | −5.04*** | −5.35*** | −14.94*** |
| | (0.22) | (0.32) | (0.16) | (−1.71) | (−3.08) | (−2.57) | (−2.77) | (−2.94) |
| Controls | NO | NO | NO | YES | NO | NO | NO | YES |
| Obs. | 1,056 | 1,096 | 1,056 | 1,044 | 1,056 | 1,096 | 1,056 | 1,044 |
| $R^2_{adj}$ | −0.1% | −0.1% | −0.1% | 3.0% | −0.0% | −0.0% | −0.1% | 0.0% |

| | Credits | | | | Commodity | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $S_l$ | 3.23 | | | | | |
| | (0.48) | | | | | |
| $S_g$ | | 13.06 | 13.08 | 26.04 | 14.37 | 19.33 |
| | | (0.79) | (0.78) | (1.09) | (0.72) | (0.88) |
| $S_{l-g}$ | | | −1.73 | 4.25 | | |
| | | | (−0.20) | (0.48) | | |
| C | 5.25* | 5.37* | 5.40* | 3.20 | −6.36* | −17.18 |
| | (1.75) | (1.82) | (1.80) | (0.19) | (−1.91) | (−0.62) |
| Controls | NO | NO | NO | YES | NO | YES |
| Obs. | 824 | 824 | 824 | 780 | 274 | 268 |
| $R^2_{adj}$ | −0.1% | −0.0% | −0.2% | 0.3% | −0.2% | −1.00% |

Table A.9: **Inflation surprises and asset skewness.** This table summarizes the impact of macroeconomic inflation surprises on asset skewness. We regress one-month ahead realized skewness on end-of-month inflation surprises for different asset classes. For the remainder, regression specifications and definitions follow table 5. Shown are the predictive regression estimates, its corresponding t-values, the number of observations (Obs.), and the adjusted $R^2$. The sample period runs from 31-03-1997 until 31-12-2019. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level

| | Equity | | | | Bonds | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $S_l$ | 4.17 | | | | −3.82 | | | |
| | (1.60) | | | | (−1.05) | | | |
| $S_g$ | | 3.78 | 2.65 | 9.42 | | −7.21 | −8.08 | −8.35 |
| | | (0.58) | (0.47) | (1.39) | | (−1.42) | (−1.49) | (−1.49) |
| $S_{l-g}$ | | | 5.14 | 4.28 | | | −1.07 | −0.36 |
| | | | (3.08) | (1.79) | | | (−0.43) | (−0.14) |
| C | 0.77 | 1.00 | 0.77 | −7.38 | −5.15*** | −4.98*** | −5.18*** | −10.85 |
| | (0.29) | (0.40) | (0.27) | (−1.22) | (−2.82) | (−2.83) | (−2.85) | (−1.55) |
| Controls | NO | NO | NO | YES | NO | NO | NO | YES |
| Obs. | 1,032 | 1,088 | 1,032 | 982 | 1,032 | 1,088 | 1,032 | 982 |
| $R^2_{adj}$ | 0.1% | −0.0% | 0.0% | 1.6% | 0.1% | 0.1% | 0.1% | 0.2% |
| | Credits | | | | Commodities | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | | |
| $S_l$ | 3.05 | | | | | | | |
| | (0.56) | | | | | | | |
| $S_g$ | | 3.91 | 4.30 | 10.12 | 12.16 | 16.92 | | |
| | | (0.55) | (0.60) | (1.31) | (1.33) | (1.61) | | |
| $S_{l-g}$ | | | 2.03 | 3.12 | | | | |
| | | | (0.27) | (0.39) | | | | |
| C | 5.25* | 5.32* | 5.28* | 8.52 | −6.51** | −18.12 | | |
| | (1.80) | (1.80) | (1.80) | (0.72) | (−2.04) | (−1.17) | | |
| Controls | NO | NO | NO | YES | NO | YES | | |
| Obs. | 824 | 824 | 824 | 780 | 272 | 255 | | |
| $R^2_{adj}$ | −0.1% | −0.1% | −0.2% | −0.1% | 0.3% | 0.1% | | |

Table A.10: **Negative expected return tests: local macroeconomic growth surprises.** This table summarizes the results of multiple tests regarding negative expected returns and local growth surprises. Panel A reports the number (#) and fraction (%) of expected return forecasts for which the 90% confidence interval lies below zero. Panel B reports the estimate of the non-linearity regression using local growth surprises. $D$ takes value one if expected excess returns are negative, and zero otherwise. The last row report the Wald test on being statistically significantly different from zero ($\chi^2(1)$). The p-values are computed with double-clustered (by time and asset) standard errors. Panel C reports the minimum expected return and its p-value based (within parenthesis) on the minimum expected return test of Eleswarapu and Thompson (2007) using 10.000 bootstraps. The sample period runs from 31-03-1997 until 31-12-2019. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| **Panel A** | Equity | Bonds | Credits | Commodities |
|---|---|---|---|---|
| $E(r) < 0$ | 93 of 1,056 | 0 of 1,056 | 31 of 824 | 93 of 274 |
| | (8.8%) | (0.0%) | (3.8%) | (33.9%) |
| **Panel B** | | | | |
| $\alpha$ | 0.67*** | 0.03*** | 0.33** | 0.51 |
| | (2.77) | (3.34) | (2.17) | (1.56) |
| $G_{sur}$ | 4.84** | -0.11 | 0.33 | -0.08 |
| | (2.47) | (-1.37) | (0.77) | (-0.04) |
| $D \times G_{sur}$ | 5.00 | - | 1.15 | 5.84** |
| | (1.62) | - | (1.59) | (2.23) |
| $G_{sur} + D \times G_{sur}$ | 9.83*** | - | 1.48*** | 5.76*** |
| | (23.54) | - | (13.48) | (9.19) |
| **Panel C** | | | | |
| Min. $E(r)(\%)$ | -2.45*** | 0.00 | -1.03** | -4.44*** |
| | (0.00) | (0.99) | (0.03) | (0.00) |

Table A.11: **Negative expected return tests: local macroeconomic inflation surprises.** This table summarizes the results of multiple tests regarding negative expected returns and local inflation surprises. Panel A reports the number (#) and fraction (%) of expected return forecasts for which the 90% confidence interval lies below zero. Panel B reports the estimate of the non-linearity regression using local inflation surprises. $D$ takes value one if expected excess returns are negative, and zero otherwise. The last row report the Wald test on being statistically significantly different from zero ($\chi^2(1)$). The p-values are computed with double-clustered (by time and asset) standard errors. Panel C reports the minimum expected return and its p-value based (within parenthesis) on the minimum expected return test of Eleswarapu and Thompson (2007) using 10.000 bootstraps. The sample period runs from 31-03-1997 until 31-12-2019. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| **Panel A** | Equity | Bonds | Credits | Commodities |
|---|---|---|---|---|
| $E(r) < 0$ | 6 of 1032 | 0 of 1032 | 0 of 822 | 0 of 272 |
| | (0.6%) | (0.0%) | (0.0%) | (0.0%) |
| **Panel B** | | | | |
| $\alpha$ | 0.50** | 0.03*** | 0.27** | -0.13 |
| | (2.26) | (3.42) | (2.06) | (-0.47) |
| $I_{sur}$ | 1.78*** | -0.04 | 0.94** | 0.44 |
| | (2.82) | (-0.77) | (2.03) | (0.81) |
| $D \times G_{sur}$ | 2.79 | 0.10 | - | 3.42** |
| | (1.37) | (1.05) | - | (2.03) |
| $G_{sur} + D \times G_{sur}$ | 4.58*** | 0.06 | - | 3.87*** |
| | (11.09) | (1.43) | - | (5.56) |
| **Panel C** | | | | |
| Min. $E(r)(\%)$ | -4.24*** | -0.02 | -1.03*** | -1.31 |
| | (0.00) | (0.22) | (0.01) | (0.12) |

Table A.12: **Return predictability and economic surprise momentum: local surprises.** The table summarizes the results of predictive panel regressions split in subsamples based on the sign of current and future local surprises. Regression specifications follow table 4. "Equal Sign" ("Different Sign") denotes the sample for which the current local surprise index has (not) the same sign as next month (Panel A and B). "Equal Own Sign" ("Different Own Sign") denotes the sample for which the weighted individual surprise has (not) the same sign as next month (Panel C and D). "Equal Cross Sign" ("Different Cross Sign") denotes the sample for which the weighted cross terms have (not) the same sign as next month (Panel C and D). Shown are the predictive regression estimates, its corresponding t-values, the number of observations (Obs.), and the adjusted $R^2$. The sample period runs from 31-03-1997 until 31-12-2019. The t-values, shown between parenthesis, are adjusted for time and asset clustering, except for commodities where we used Newey-West corrected t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| **Panel A: Growth** | Equity | | Bonds | | Credits | |
|---|---|---|---|---|---|---|
| | $S_l$ | $R^2$ | $S_l$ | $R^2$ | $S_l$ | |
| Equal Sign | 2.58$^{***}$ | 2.0% | -0.05 | 0.4% | 1.37$^{**}$ | 4.6% |
| | (3.06) | | (-0.77) | | (2.23) | |
| Different Sign | 0.85$^{**}$ | -0.1% | 0.01 | -0.3% | -0.36 | -0.1% |
| | (2.16) | | (0.26) | | (-0.53) | |
| Full Sample | 2.25$^{***}$ | 1.5% | -0.04 | 0.1% | 0.95$^{**}$ | 1.9% |
| | (3.11) | | (-0.70) | | (2.22) | |
| **Panel B: Inflation** | Equity | | Bonds | | Credits | |
| | $S_l$ | $R^2$ | $S_l$ | $R^2$ | $S_l$ | $R^2$ |
| Equal Sign | 0.74$^{**}$ | 0.4% | -0.04$^{**}$ | 0.9% | 0.04 | -0.2% |
| | (2.30) | | (-2.07) | | (0.18) | |
| Different Sign | 0.65$^{**}$ | 0.5% | -0.01 | -0.0% | -0.04 | -0.2% |
| | (2.33) | | (-1.17) | | (-0.23) | |
| Full Sample | 0.68$^{***}$ | 0.5% | -0.03$^{**}$ | 0.5% | 0.01 | -0.1% |
| | (2.94) | | (-2.28) | | (0.05) | |
| **Panel C: Growth** | Equity | | Bonds | | Credits | |
| | $S_l$ | $R^2$ | $S_l$ | $R^2$ | $S_l$ | $R^2$ |
| Equal Own Sign | 2.44$^{**}$ | 1.9% | 0.04 | 0.2% | 1.19$^{**}$ | 3.1% |
| | (2.64) | | (-0.63) | | (2.38) | |
| Different Own Sign | 0.89 | -0.2% | -0.01 | -0.4% | -0.17 | -0.4% |
| | (1.60) | | (-0.29) | | (-0.98) | |
| Equal Cross Sign | 2.48$^{**}$ | 2.1% | -0.06 | 0.6% | 1.22$^{**}$ | 4.0% |
| | (2.32) | | (-0.88) | | (2.22) | |
| Different Cross Sign | 1.64$^{***}$ | 0.3% | 0.03 | -0.1% | 0.02 | -0.3% |
| | (2.73) | | (0.65) | | (0.04) | |
| **Panel D: Inflation** | Equity | | Bonds | | Credits | |
| | $S_l$ | $R^2$ | $S_l$ | $R^2$ | $S_l$ | $R^2$ |
| Equal Own Sign | 1.03$^{***}$ | 1.2% | 0.01 | -0.2% | -0.05 | -0.3% |
| | (3.05) | | (0.43) | | (-0.27) | |
| Different Own Sign | 0.37 | 0.0% | -0.05$^{***}$ | 2.1% | 0.02 | -0.2% |
| | (1.23) | | (-3.59) | | (0.12) | |
| Equal Cross Sign | 0.83$^{*}$ | 0.6% | 0.00 | -0.2% | -0.08 | -0.2% |
| | (1.94) | | (0.02) | | (-0.35) | |
| Different Cross Sign | 0.48$^{*}$ | 0.2% | -0.05$^{***}$ | 1.9% | 0.01 | -0.2% |
| | (1.93) | | (-2.92) | | (0.03) | |

# Chapter 3

# Non-Standard Errors in Asset Pricing: Mind Your Sorts[1]

## 3.1 Introduction

Characteristic-based portfolio sorting is a widely used procedure in modern empirical finance. Researchers deploy the procedure to test theories in asset pricing, to study a wide range of pricing anomalies, and to identify profitable investment strategies. Based on this procedure, the academic literature in finance documents a range of factors that appear relevant for the cross-section of equity returns, known as the "factor zoo" (Cochrane, 2011). As researchers face a number of design choices when engaging in portfolio sorting, the exact construction procedure is not uniform across studies. Potentially, the differential design choices lead to considerable variation in outcomes. Menkveld et al. (2023) refer to such variation in outcomes due to choices in the evidence-generating process as non-standard errors.

In this paper, we study the extent to which the differential design choices in portfolio sorting matter for factors, factor models, and non-standard errors. A better understanding of the design choices that matter allows researchers to more effectively show the robustness of their findings and to reduce non-standard errors in future work, while also facilitating model selection exercises and helping interested readers in interpreting presented results.

We consider eleven construction choices that researchers face in their research design. These choices are: (1) 70/30 or 80/20 breakpoints, (2) NYSE or NYSE-AMEX-Nasdaq (NAN) breakpoints, (3) including or excluding firms with a negative book equity value, (4) including or excluding microcaps, (5) imposing a price filter or not, (6) including or excluding utility firms, (7) including or excluding financial firms, (8) industry neutralization or not, (9) value-weighting or equal-weighting, (10)

[1]This chapter is based on Soebhag, Van Vliet, and Verwijmeren (2022).

independent or dependent sorts, and (11) sorting on the most recent market capitalization or from June. We construct factors using each possible combination of choices, which leads to 2048 ($2^{11}$) construction combinations.

Our analysis centers on maximum Sharpe ratios as these allow us to assess both individual factors and factor models (Barillas and Shanken (2017), Fama and French (2018)). Based on data on U.S. stock returns from January 1972 to December 2021, we find that factors exhibit large variation in Sharpe ratios within our set of possible construction methods. As an illustration, figure 1 shows the gross annualized Sharpe ratio of the canonical value factor (HML) of Fama and French (1993) for the 2048 possible construction choices. The median Sharpe ratio across the choice set is 0.49. The figure shows that the variation in obtained annualized Sharpe ratios is substantial. Depending on how we create the HML factor, Sharpe ratios vary between 0.15 and 1.24. Our paper shows that the same design choices can also strongly affect the Sharpe ratios of other factors.

The non-standard errors in our setting can be defined as the standard deviation of the generated Sharpe ratios across the possible construction methods. We find that these non-standard errors are sizable relative to standard errors, across all factors. In multiple cases, the non-standard errors exceed the standard errors. For example, the non-standard error for the post-earnings announcement drift factor is 0.10, whereas the standard error ranges between 0.04 and 0.09. The average ratio of the non-standard error to the standard error across factors is 1.18. As such, factor returns are not only a function of their sorting characteristic, but also a function of their construction choices.

The above non-standard errors are based on researchers randomly choosing construction methods, which helps in assessing the room provided by these choices to optimize along a given criteria, given that researchers may have incentives to engage in p-hacking (Harvey, 2017). An alternative calculation of non-standard errors takes into account that not all choices are equally likely, as researchers could have good reasons to select a particular choice (some choices might be more "reasonable"), or researchers are simply more likely to select a choice that they have seen more often in earlier work. For our alternative non-standard error measure, we classify the choices made in a set of 323 empirical asset pricing papers. We exploit the popularity of each binary choice in earlier work to construct non-standard errors that take these probabilities into account. Interestingly, no binary option is so dominant that it represents close to 100% of the observed choices, and some options are even selected with a roughly 50% probability. Using these probabilities, we find that non-standard errors remain sizable, with an average ratio of non-standard errors to standard errors of 1.08.

Factor models have been compared against each other based on factor construction methodology choices of a single paper. Our paper aims to perform a model comparison without relying on a single set of construction choices but by considering a wide set of potential choices. Barillas and Shanken (2017) show that for models with traded factors, the extent to which each model is able to price factors in the other model is what matters for model comparison, not the test assets. They propose the use of maximum squared Sharpe ratios as a model comparison metric, which Fama and French (2018) use to evaluate their 3-factor, 5-factor, and 6-factor models.[2] We find that on average the factor models of Barillas and Shanken (2018) and Daniel, Hirshleifer, and Sun (2020) have the largest maximum Sharpe ratio. Importantly, this maximum Sharpe ratio, and with it the dominant factor model, also varies across construction methods.

We further find a large discrepancy in optimal mean-variance weights within factor models. Moreover, our findings indicate that economic significance, i.e., how much gain could be realized by a mean-variance investor, is sensitive to construction methods. In additional tests, we study whether variation in construction choices affects factor exposure, liquidity and transaction costs of a portfolio. We again find that portfolio construction methods matter. For example, equal-weighting leads to portfolios with higher illiquidity than value-weighting, which consequently results in higher risk-adjusted gross returns.

Overall, we conclude that factor design choices matter. Our results imply that multiple construction methods should be considered to reduce the potential for data mining. We find that particularly important choices are those concerning the use of NYSE or NAN breakpoints, including or excluding micro stocks, using industry-adjusted characteristics or not, and equal- versus value-weighting. For future studies, one way forward is to consider these choices in a "specification check" (Brodeur et al. (2020), Mitton (2022)), in which the distribution of the results from the combinations of these methodological possibilities are reported.

Another way forward is for studies to be more uniform in their choices. Based on our analysis, the conservative choices of using NYSE breakpoints, excluding microcaps, and using value-weighting reduces the average non-standard error by 70%. This relatively large reduction in the average non-standard error is an important contribution for two reasons.

The first reason why singling out the most important choices matters is that researchers are still not uniform in their choices. Our analysis links to earlier work by Hou et al. (2019), which shows that the performance of factors is sensitive to the breakpoints being used, and Hou et al. (2020), which shows that many anomalies disappear when microcaps are excluded. Our examination of the combined set of construction choices leads us to conclude that the conservative choices would be to use NYSE breakpoints and to exclude microcaps. However, when we consider the 25 publications in the Journal of Finance, Review of Financial Studies, and Journal of Financial Economics in 2021 and 2022 that employ portfolio sorting, only four

[2] Barillas et al. (2020) compare a range of models using the maximum squared Sharpe ratio and find that a variant of the Fama and French (2018) 6-factor model, with a monthly updated version of the value factor, emerges as the dominant model.

of these papers use NYSE breakpoints rather than NAN breakpoints and only six papers exclude microcaps.

The second reason why the large observed reduction in non-standard errors matters is that our results imply that by just keeping three out of the eleven choices fixed, empirical asset pricing researchers can greatly facilitate the interpretability of their presented results. In other words, our evidence that a range of design choices does not have very large effects is also a meaningful contribution, especially given the concerns around p-hacking typically present in the field. As such, part of the conclusion of this paper provides a comforting message to the empirical asset pricing field.

Our paper relates to empirical studies on the replicability of market anomalies.[3] Our analysis centers on potential drivers of the variation in outcomes. Our setup links to other recent papers studying design choices in asset pricing, particularly on the value premium. Kessler et al. (2020) take a practitioner's perspective to show the impact of design choices on the value premium on the S&P 500, while Hasler (2022a) shows that alternative choices lead to a value premium that is smaller than originally thought. More generally, Hasler (2022b) concludes that statistical biases from research decisions can explain around a fifth of the return predictability in the literature. Our goal is to get an idea about the magnitude of non-standard errors by assessing the importance of a range of construction choices, and we additionally aim to compare factor models on an "apples-to-apples" basis.

Our work contributes to the non-standard error literature. Non-standard errors are introduced by Menkveld et al. (2023), who argue that a layer of uncertainty in academic work is due to the evidence-generating process that exposes variation across choices by researchers, in addition to the traditional standard errors following from the uncertainty in sample estimates of population parameters (the data-generating process). By letting 164 research teams independently test the same market microstructure hypotheses on the same sample of trade records, they find that non-standard errors are sizable. Our paper can be thought of as modelling $N$ hypothetical researchers who construct factor returns. Our approach allows us to examine non-standard errors both when researchers freely exploit the variation of choices available to them, and when researchers base their choices on conventions in

[3]McLean and Pontiff (2016) test anomalies out-of-sample and find that the performance of identified anomalies diminishes after publication. Harvey et al. (2016) derive threshold levels to take into account potential data mining. Based on a multiple testing framework, they find that many anomalies are likely false discoveries. Linnainmaa and Roberts (2018) find that a similar conclusion can be drawn when examining pre-sample periods. Hou et al. (2020) test 452 anomalies by using a single factor construction procedure. They find that around two-thirds of the anomalies fail to replicate, even if they do not adjust for multiple hypothesis testing. On the other hand, Yan and Zheng (2017) use a bootstrap approach to evaluate fundamental-based anomalies and find that many fundamental signals are significant predictors of cross-sectional stock returns, even after accounting for data mining. Chen and Zimmermann (2022) and Jensen et al. (2022) show that they are able to successfully reproduce the majority of asset pricing factors.

earlier work.[4] We find relatively high non-standard errors in both scenarios.

The remainder of this paper is organized as follows. We describe the data and factor models in section 3.2. Section 3.3 describes the empirical variation in sorting methods. In section 3.4 we examine the importance of factor construction choices for Sharpe ratios and calculate non-standard errors. Section 3.5 examines whether factor construction choices impact model selection exercises. Section 3.6 shows how different construction methods affects several key portfolio characteristics. Section 3.7 concludes.

## 3.2 Constructing Factor Models

We obtain monthly returns and prices for U.S. equities from the Center for Research in Security Prices (CRSP). Accounting information is retrieved from the Compustat Annual and Quarterly Fundamental Files. Our sample consists of stocks listed on the NYSE, AMEX, and Nasdaq and with share codes 10 or 11, which limits our sample to common stocks. The sample period spans January 1972 to December 2021, thereby covering 600 months of factor returns.[5]

We use multiple factor models, originating from Fama and French (2015), Hou et al. (2015), Fama and French (2018), Barillas and Shanken (2018), and Daniel, Hirshleifer, and Sun (2020). Table 1 summarizes the factors underlying the factor models and their key construction choices as used in their original studies.[6] The market factor is a part of all models. The Fama-French 5-factor model of Fama and French (2015) (FF5) consists of the market, size (SMB), value (HML), profitability (RMW) and investment (CMA) factors. The factors are constructed, originally, by using a 2 by 3 independent sort between size and the characteristic. The size sort uses a median breakpoint, and the sorting characteristic is split by the 30th and 70th percentile, both on the NYSE universe. All factors of the Fama-French 5 factor model are rebalanced yearly. The 6-factor model of Fama and French augments the 5-factor model by adding the momentum (UMD) factor, with the resulting model being abbreviated as FF6. The UMD factor differs only in the rebalancing, which is monthly. In addition, we construct a cash-based version of the $RMW$ factor (named $RMW(CP)$) for both models as suggested by Fama and French (2018). This results in models that we abbreviate as $\text{FF5}_c$ and $\text{FF6}_c$. The Q factor model of Hou et al. (2015) consists of the market factor, size factor, investment (IA) factor, and return on equity (ROE) factor. In the original set-up, these factors are derived from a 2x3x3 independent sort.[7] Barillas and Shanken (2018) combine factors from the FF models

[4] Walter et al. (2022) complements our paper by focusing on the impact of individual decision nodes and their economic interpretation.

[5] The starting year is 1972 as we require quarterly earnings announcements dates (to construct the price earnings announcement drift factor) and quarterly book equity data (to construct the return on equity factor).

[6] Definitions of the sorting variables are provided in Appendix 3.10.

[7] We examine the 2x3x3 versus the 2x3 sorting procedure in Appendix 3.11.

and Q model into a six-factor model (BS), consisting of the market factor, size factor, a monthly-updated value factor, the momentum factor, the growth in book factor, and the return on equity factor. Daniel, Hirshleifer, and Sun (2020) (DHS) construct a three-factor model consisting of the market factor, financing (FIN) factor, and the post-earnings announcement drift (PEAD) factor. Both the FIN and PEAD factor use 20-80 breakpoints in the characteristic dimension. The PEAD factor is rebalanced monthly.

We construct factor portfolios by sorting on both market capitalization and a factor characteristic. The size dimension is split into a "Small" and a "Big" segment based on the median. The characteristic dimension is split into a "Low", "Neutral", and "High" portfolio based on two breakpoints.[8] This procedure, the 2x3 sorting, results into six portfolios: Small.Low, Small.Neutral, Small.High, Big.Low, Big.Neutral and Big.High. We create the factor portfolio by taking a long position in the Small.High and the Big.High portfolio and a short position in the Small.Low and Big.Low portfolio:

$$Factor = (Small.High + Big.High)/2 - (Small.Low + Big.Low)/2 \tag{3.1}$$

## 3.3 Variation in Sorting Methods

Researchers face a large number of methodological decisions when testing hypotheses. To examine the methodological choices that have been made in the empirical asset pricing literature focusing on portfolio sorting, we survey 323 empirical articles in the top finance journals between 1965 and 2018, based on the list of papers that Harvey and Liu (2019) constructed for their census of the "factor zoo". In this section, we start by describing common portfolio sorting decisions that researchers make in their studies. We then document how often these research choices are being made.

### 3.3.1 Construction choices

Based on the data and methodology sections of the 323 empirical asset pricing studies, we find that there are eleven construction choices commonly being mentioned.[9] These eleven choices result in a set of 2048 ($2^{11}$) construction choices, which translates to 2048 different versions for each factor and factor model. In this subsection, we explain the eleven choices one by one.

[8] For the financing factor we follow the approach in Daniel, Hirshleifer, and Sun (2020) by separately sorting all repurchasing firms into two groups using a median breakpoint, and sorting all issuing firms into three groups using two breakpoints.

[9] We do not consider the sample period as a construction choice, as the convention is to start in the year when all relevant data become available and finish in the most recent year with full data availability (at the time of the analysis). The download date of the data can matter: Akey et al. (2022) show that Fama-French factor returns vary across different factor vintages.

#### Characteristic breakpoints

Common practice in the academic finance literature has been to create portfolios by sorting on characteristics positively associated with expected returns. Various breakpoints have been proposed to create long-short portfolios. One standard procedure is to construct factors using a 2×3 sorting procedure as in Fama and French (1993). First, stocks are sorted by their market capitalization, whereby stocks are split into "small" and "big" classifications based on the NYSE median break-point. Second, and independently, stocks are sorted on their characteristic, whereby stocks are classified into "high" and "low" based on the 30th and 70th percentile (calculated over the NYSE universe) of the characteristic. The intersection of these classifications results into six portfolios, from which the high-minus-low portfolio is derived.

The 30th and 70th percentile breakpoints are thus one popular choice, used in, for example, Fama-French models (Fama and French (2018)) and in the Q factor model (Hou et al. (2015)).[10] However, many others have chosen to deploy the 20th and 80th percentile to sort portfolios in the characteristic dimension. Examples of studies using this method are McLean and Pontiff (2016), Stambaugh and Yuan (2017) and Daniel, Hirshleifer, and Sun (2020). The consequence of using the latter choice is that stocks with more extreme characteristics are selected into portfolios. We construct different versions of factors where we either use the 30th-70th breakpoint or the 20th-80th breakpoint in the characteristic dimension.[11]

#### Breakpoints universe

A common choice is to calculate breakpoints over the NYSE universe. However, a popular alternative is to calculate breakpoints over the NYSE-AMEX-Nasdaq (NAN) universe, such as done by McLean and Pontiff (2016), Stambaugh and Yuan (2017) and Yan and Zheng (2017). Since Nasdaq and AMEX stocks have a tilt towards smaller stocks, the median market capitalization is always higher under the NYSE criteria relative to the NAN criteria. As such, using NAN breakpoints is likely to provide an overweight towards micro- and small-cap stocks relative to using NYSE breakpoints.

#### Negative book equity value

Firms with a negative book equity value were rare before 1980 (Fama & French, 1993). However, they represent a larger proportion of firms over time, even though negative book equity has no obvious interpretation due to a firm's limited liability structure (Brown et al., 2008). Many practitioners and academics omit negative book equity firms from their analysis, but an even larger set of papers still contains

[10]Several studies download portfolios from the Kenneth French data library, thereby (implicitly) choosing for the portfolio construction choices of Fama and French.

[11]Rather than using double-sorted factors to control for size effects, studies can also construct univariate sorts, with decile breakpoints. If expected returns linearly increase in those stock characteristics, such factor portfolios will tend to have higher average returns and non-standard errors would increase with decile sorts.

analyses with negative book equity value firms included. Brown et al. (2008) show that negative book equity firms are disproportionately represented in extreme growth and value sectors.

### Microcaps

We also consider the inclusion and exclusion of microcaps as a construction choice. Microcaps are typically defined as stocks with a market capitalization below the 20th percentile for NYSE stocks. Fama and French (2008) find that microcaps account for 60% of the number of stocks, but only capture 3% of the total market capitalization. In addition, they find that microcaps have the highest cross-sectional volatility of returns and show large dispersion in sorting characteristics. From a practical perspective, these small stocks are out of reach for many (institutional) investors. In addition, microcaps are more expensive to short due to high shorting fees (Drechsler and Drechsler (2014)), they may be illiquid, and they have high transaction costs (Novy-Marx and Velikov (2016)). Nevertheless, microcaps are typically included in many studies. Hou et al. (2020) find that many anomalies documented in the literature do not survive after excluding microcaps. Excluding microcaps increases the median market capitalization, reduces typical return volatility, and increases the market share of stocks below the median.

### Filtering on price

A price filter leads researchers to exclude firms solely based on absolute share prices. More specifically, stocks are dropped for having share prices below a minimum, which typically varies between $1 and $5. In fact, the most common price filters use a minimum of exactly $1 (e.g, Lee and Swaminathan (2000)) or exactly $5 (e.g, Amihud (2002)). Applying a price filter removes potentially highly illiquid and often highly volatile stocks.

### Utility firms

Utility firms typically engage in the generation, transmission and/or distribution of electricity, gas, or steam, while the category also includes firms active in waste management. In empirical corporate finance studies, it is standard to exclude utility firms from the analysis, as they are seen as different due to the regulations utility firms have to comply with. These regulations could also explain the exclusion of utility firms in asset pricing studies, such as in Hirshleifer and Jiang (2010), who argue that mispricing is more constrained among regulated industries. Still, most empirical asset pricing studies incorporate utility firms in their analysis.

### Financial firms

Excluding financial firms from the sample is not unusual in empirical studies. The argument for this exclusion criteria is that financial services are fundamentally different, resembling the potential argument for utility firms. Fama and French (1992) explicitly mention that financial firms have high leverage, which is normal for such firms, and that it probably does not have the same meaning as for non-financial firms, where high leverage is more likely to indicate distress. Still, many other papers include financial firms, such as Stambaugh and Yuan (2017). Including financial firms may especially impact factor returns when factors are not hedged against industry exposure.

### Industry hedging

Additionally, we consider industry hedging as a construction choice. The unconditional predictive power of stock characteristics may stem from their across-industries component or from their firm-specific (within-industries) component, or from both (Ehsani et al. (2021)). A consequence of unconditional sorting is that factor portfolios obtain differential exposure towards specific industries. To illustrate, constructing the unconditional value factor overweights sectors that contain stocks with high book-to-market ratios, such as utility firms in the long leg, whereas the short value leg gets excess exposure towards technology stocks.

Daniel, Mota, Rottke, and Santos (2020) suggest that sorting stocks, unconditionally, tends to pick-up unintended (industry) risks, generating portfolios that are no longer mean-variance efficient. Sector-concentrated portfolios are more volatile because stocks within the same sector are highly correlated. Under-diversification due to these exposures do not implicitly reveal information about the expected returns of factors and hedging these exposures is a choice that can be made in order to improve risk-adjusted returns.[12] A comparison of the standard and industry-hedged factors shows that industry adjustment often improves factor performance (Asness et al. (2000), Novy-Marx (2013)).

We construct industry-hedged factors, in addition to unhedged factors, by normalizing the sorting characteristic into an industry-adjusted characteristic as follows:

$$S^*_{i,t} = (S_{i,t} - \bar{S_{i,j,t}})/(S_{max,j,t} - S_{min,j,t}) \tag{3.2}$$

$S_{i,t}$ ($S^*_{i,t}$) denotes the (industry-adjusted) sorting characteristic. $\bar{S_{i,j,t}}$, $S_{max,j,t}$ and $S_{min,j,t}$ are equal to the cross-sectional mean, maximum and minimum, respectively, of the sorting characteristic $S$ for industry $j$. We use the Fama-French 12-industry classification.

### Value-weighting vs. equal-weighting

There are several weighting schemes that a researcher can select when constructing a portfolio. The literature focuses predominantly on value- or equal-weighting portfolios. Different choices regarding weights results in different portfolio compositions and consequently in differential portfolio characteristics and performance. When

[12] Especially practitioners typically add industry constraints in portfolio construction processes to avoid concentration risks.

using the value-weighting approach, these exposures depend on the size of the specific companies. The risk and return will be driven predominantly by the largest companies in the investment universe. Value-weighted portfolios typically serve as a benchmark against which portfolio managers are evaluated, highlighting the relevance of value-weighting in practice. Nevertheless, the majority of studies before 2010 use equal-weighting when constructing factor portfolios (Green et al., 2013). In robustness tests, it is common for papers to show the results when the alternative weighting choice would have been selected.

**Independent versus dependent**

Independent sorting is the most commonly used sorting procedure deployed in the literature. A major drawback is that independent sorting may result in sparse portfolios, with the consequence that a factor portfolio is not well-diversified. In some cases, independent sorting may even result in empty portfolios, which is especially an issue in international or smaller samples (Ang et al. (2006), Novy-Marx (2013), Wahal and Yavuz (2013)). Dependent sorting alleviates the problem of sparse portfolios by sequentially stratifying stocks into portfolios. However, implementing a dependent sorting procedure raises the question of what order of the sort should be used, especially when sorting on more than two factors. For the 2x3 procedure, the standard is to first sort on size, and then on the sorting characteristic, i.e., there is little degree of freedom in this choice. However, when we consider a 2x3x3 dependent sort, it is not clear what the ordering should be, allowing for a wider playing field.

**When to observe market capitalization**

Common practice is to construct size-breakpoints based on the market capitalization of firms at the end of June of the current year $t$, and update this yearly, following Fama and French (1992). Some studies have chosen to use the market capitalization in the previous month in their size sort. For example, Daniel, Hirshleifer, and Sun (2020) do so when constructing the PEAD factor, and Ang et al. (2006) in their analysis of the idiosyncratic volatility anomaly. One argument in favor of using the most recent market capitalization might be to use timely information to construct the size sort. On the other hand, this may result into more turnover, since one rebalances the size sorts each month instead of each year.

### 3.3.2 Distribution and correlation of choices

We continue our meta-analysis in this section by documenting how frequent certain choices are being made, again based on the list of 323 empirical articles constructed for Harvey and Liu (2019). We select the choices made for the main analyses in the study. For instance, if a study notes that results are robust to excluding microcaps, the main choice was to include microcaps. If choices are being made randomly, the expected proportion of studies in which a particular option is selected will be close to 50% for each of our binary choices. However, if there are good reasons for particular design choices, or if authors build on the choices being made in earlier work, then we might expect some choice options to be selected (close to) 100% of the time. We show the percentage of studies in which a particular design option is selected in Table 2.

For some design choices the distribution is relatively equal. The number of studies reporting to use 30-70 breakpoints roughly equals the number of studies reporting to use 20-80 breakpoints.[13] NYSE breakpoints are used by 41.5% of the studies reporting this information. Value-weighting returns (58.5%) is slightly more popular than equal-weighting returns (41.5%). We confirm an increased popularity of value-weighting over time in our data, which explains why equal-weighting is not as popular as it was before 2010 (Green et al., 2013).

No design option is selected in 100% of the cases. The most popular design option is to include utilities, as utility firms are excluded from the sample in only 9.9% of the relevant studies. Financial firms are excluded in 28.8% of the cases. Other popular options are to not impose industry neutrality (88.5%), to include microcaps (88.2%), to not impose a price filter (81.7%), and to include firms with negative book equity values (78.0%). We further find that the proportion of studies using independent sorts is 71.8% and that in 67.4% of the studies the market capitalization from last June is used.

The percentages reported in table 2 indicate that no design option in our set is extremely rare. Regardless of the choice being made, a researcher can always cite at least ten other papers making the same choice. To examine whether there are combinations of choice options that are particularly rare, we report the correlation matrix of the eleven choices in table 3.

The typical correlation coefficient is not particularly high. If one, for example, includes microcaps, there is an increased probability of using value-weighting rather than equal-weighting (correlation coefficient of 0.22), but the correlation is not so strong that a choice for equal-weighting would be considered as exceptional. The highest correlations are observed between using independent sorting and using the size from June (0.54) and between excluding firms with negative book equity values and including financial firms (-0.53).

## 3.4 The Impact of Construction Choices and the Size of Non-Standard Errors

In this section, we examine the impact of portfolio design choices on Sharpe ratios. We focus on Sharpe ratios as these allow us to assess both individual factors in this section, and factor models in the next section (Barillas and Shanken (2017), Fama

[13] We classify studies that use quintile sorts as using 20-80 breakpoints and studies that use tercile breakpoints as using 30-70 breakpoints.

and French (2018)). In addition, in this section we compute non-standard errors and compare these with estimated standard errors. This section concludes with an analysis of potential reductions in non-standard errors.

### 3.4.1 Construction choices and Sharpe ratios

Table 4 reports summary statistics of the factors that we include in our sample, based on the factor models from table 1. These factors are the size (SMB), value (HML and the monthly version, HML(m)), operating-based profitability (RMW), cash-based profitability (RMW(cp)), investment (IA and CMA), momentum (UMD), return on equity (ROE), financing (FIN), and post-earnings announcement drift (PEAD) factor. The table shows the annualized average return and Sharpe ratio per factor, both value-weighted and equally-weighted, when averaged over the set of construction methods. Value-weighted factor returns range between 1.91% (SMB) and 8.00% (UMD) per year, with Sharpe ratios ranging between 0.18 (SMB) and 1.10 (PEAD). Returns and Sharpe ratios for equal-weighted factors tend to be higher, except for the size factor.

Figure 2 shows the Sharpe ratio distribution across sets of construction choices for each factor, based on long-short factor returns. We construct a factor 2048 times by using the 2048 different factor construction methods. The cutoff for the potential price filter is set at $5. Figure 2A shows the distribution of value-weighted Sharpe ratios and Figure 2B shows the distribution of equal-weighted Sharpe ratios. Both figures show differences in median Sharpe ratios across factors, but also substantial variation in Sharpe ratios within a factor.[14] For example, based on value-weighting, the Sharpe ratio of the CMA factor ranges between 0.18 and 0.90, the Sharpe ratio of the UMD factor ranges between 0.37 and 0.78, and the Sharpe ratio of the ROE factor ranges between 0.46 and 1.06. Based on equal-weighting, the ranges are between 0.34 and 1.44 for the CMA factor, between 0.28 and 0.92 for the UMD factor, and between 0.42 and 1.40 for the ROE factor. Hence, in relative terms, the Sharpe ratio can more than double depending on design choices, and this applies to the far majority of factors. In absolute terms, the PEAD factor shows the largest variation in absolute terms, ranging from 0.73 to 1.76 for value-weighted Sharpe ratios and from 0.67 to 2.18 for equal-weighted Sharpe ratios. Overall, these results imply that construction choices matter.

We next examine how specific construction choices, in isolation, affect maximum Sharpe ratio estimates. Figure 3 shows annualized maximum Sharpe ratios by construction choice, averaged over factor models. We first vary the breakpoints that are used to classify high and low characteristics. The first two bars on the left-hand side use the 20th-80th percentile (white bar) or the 30th-70th percentile (dashed bar). The latter case yields an average annualized Sharpe ratio of 0.63, whereas 20-80 breakpoints yield a Sharpe ratio of 0.65. Intuitively, if expected returns are mono-

[14]Appendix 3.12 reports all analyses using net returns. In Appendix 3.12.3, we show that Sharpe ratio variation is also sizable after correcting for transaction costs.

tonically related to a given stock characteristic, then taking positions in stocks with more extreme characteristics would naturally result into higher returns and Sharpe ratios.

Using NAN breakpoints instead of NYSE breakpoints improves Sharpe ratios from 0.55 to 0.73, which is the largest increase within our set of choices. This choice thus comes out as important, where NYSE breakpoints represent the conservative choice. Anoter important choice is the choice whether to include microcaps or not. Including microcaps improves the average Sharpe ratio from 0.59 to 0.68, which makes excluding these firms the conservative choice.

Choices that do not lead to substantially different average Sharpe ratios include choices related to negative book equity firms, a five dollar price filter, and utility firms. Including financial firms increases the Sharpe ratio, on average, from 0.62 to 0.66. It can further be seen that eliminating industry exposures from factor returns substantially increases Sharpe ratios, which is in line with Daniel et al. (2020).

Equal-weighting portfolios improves the Sharpe ratio compared to value-weighting portfolios from 0.57 to 0.71, on average, and also comes out as one of the more important design choices. The Sharpe ratios for independent and dependent sorts are approximately similar. Finally, using the most recent market capitalization to construct factors increases the Sharpe ratio from 0.63 to 0.65 relative to using the market cap in June. Overall, our findings imply that construction choices can materially affect factor performance, especially those concerning NYSE breakpoints, micro stocks, industry-adjusted characteristics, and value-weighting.

### 3.4.2 Non-standard errors versus standard errors

Based on the above analyses, non-standard errors might be sizable. Traditionally, the focus of the empirical finance literature has been on standard errors, resulting from a data-generating process drawing samples from a population. That is, sampling uncertainty leads to standard errors when estimating population parameters, such as the mean and volatility of returns. Non-standard errors result from an evidence-generating process, which translates the sample into evidence, and which adds an additional layer of error (Menkveld et al. (2023)).

We initially model non-standard errors as the cross-sectional standard deviation across hypothetical researchers who all use different sets of construction choices. We thus obtain one non-standard error per factor, equal to the standard deviation of the 2048 different Sharpe ratios for that factor. To compare non-standard errors with standard errors, we estimate standard errors by block-bootstrapping each factor's return for a given set of construction choices. The standard error is the standard deviation of the Sharpe ratio obtained from block-bootstrapping a factor, and we block-bootstrap each series 10.000 times. Subsequently, we average the standard errors for each factor across all choices. We show the results in figure 4. The white bars indicate the non-standard error for each factor and the dashed bars denote the estimated standard errors. Besides the average standard errors, we also plot the minimum and maximum standard errors.

We find that non-standard errors are sizable relative to standard errors, across all factors. In 6 out of 11 factors, we find that the non-standard error is larger than the standard error. These factors are HML, HML(m), CMA, IA, ROE and PEAD. The non-standard errors are relatively low for SMB. The non-standard error is highest for the PEAD factor (i.e., 0.10, whereas the standard error ranges between 0.04 and 0.09). In terms of proportions (non-standard error divided by average standard error), we find that this proportion ranges between 58% (SMB) and 190% (PEAD), with the average being 118%. Overall, we conclude that non-standard errors are sizable in comparison with standard errors. The average non-standard error to standard error ratio of 118% is also relevant in comparison to the ratio of 160% found by Menkveld et al. (2023), based on a relatively high degree of researcher discretion in their experiment.

The above estimation of non-standard errors hinges on the assumption that each design choice is made with an equal probability. Consequently, each possible combination occurs once out of 2048 times in our sample. From table 2 we know that various choices do not occur with an equal probability in the literature. We can use these observed probabilities in estimating an alternative non-standard error, which we call the weighted non-standard error. More precisely, we use the implied probabilities to compute the total probability that an outcome of choices would occur. This total probability is computed by multiplying the individual implied probabilities. Subsequently, we compute the non-standard error as the weighted standard deviation across all construction choices, whereby we weight the observation by its total implied probability. We also use this weighting when calculating standard errors. We plot the weighted standard and non-standard errors in figure 5 for each factor.

We find that the typical non-standard error remains sizable. The weighted non-standard errors exceed the estimated average weighted standard error in 5 out of 11 factors. Compared to figure 4, the non-standard error of the ROE factor now falls below the standard error for that factor. The non-standard errors of the HML and PEAD factor are also slightly reduced compared to the non-weighted analysis. Overall, though, the results shown in figure 5 and figure 4 are very similar. The average weighted non-standard error divided by the average weighted standard error now ranges between 62% (FIN) and 189% (CMA), with the average being 108%.

### 3.4.3 Reducing non-standard errors

Variation in design choices allows researchers to customize samples and empirical tests to tackle specific research questions. For example, researchers might be particularly interested in patterns within financial firms, or within a particular other type of firm. Allowing some variation could thus be optimal, also to reduce the chance of missed discoveries. However, allowing for too many degrees of freedom regarding design choices and the resultant high non-standard errors severely complicates the interpretability of the results by the average reader, while also potentially inducing excessive reporting of statistically significant results.

In this section, we take this tradeoff into account and examine whether a limited set of restrictions could substantially reduce non-standard errors. This analysis follows from figure 3, which provides insights into which of the eleven design choices appear most relevant for non-standard errors. We construct two sets of potential restrictions and compare the resultant non-standard errors with those of the setting when all choices are free.

Let "Set 1" be the base case where researchers can make all of the eleven choices identified in section 3.3.2. In "Set 2" we exclude three choices that appear particularly important in figure 3: NAN breakpoints, including microcaps, and equal-weighting. Excluding these three choices could substantially reduce uncertainty in interpreting reported results. In addition, the choices can be relatively easily justified based on economic arguments. Although approximately 60% of the stocks in the CRSP sample can be considered as microcaps, they only represent about 3% of the total market capitalization of the CRSP universe. Transaction costs for microcaps are high and liquidity is low, which makes this segment of the market difficult for investors. The other choices link to microcaps. When researchers opt for equal-weighting portfolios, microcaps (and small caps) become relatively important, which tends to bias the mean return upward. This bias is limited when value-weighted returns are computed. Using NAN breakpoints also favors micro- and small caps, leading to similarly inflated anomaly profits. As such, Set 2 can be justified and resembles the choices made by e.g. Hou et al. (2020). Excluding these three choices leaves researchers with eight remaining design choices, or 256 possible combinations.[15]

A fourth choice that seems important for non-standard errors is industry neutralization. Here the trade-off might be especially important. Figure 3 suggests that the conservative choice is to not use industry-adjusted characteristics. However, Daniel, Mota, Rottke and Santos (2020) suggest that this tends to pick-up unintended (industry) risks, generating portfolios that are no longer mean-variance efficient. Hedging this exposure is thus a choice that might be sensible. However, this choice could depend on the particular research question one is after, and the unhedged approach is the more popular approach, as shown by the results in table 2.

[15] In Figure A.5, we decompose Set 2 to assess to what extent non-standard errors decrease when we impose one or two combinations of "Set 2" as a restriction. The reduction in non-standard errors is especially large when at least two of the three choices are restricted.

Instead, in "Set 3", we additionally restrict four choices that are motivated by Fama and French (1992) and Fama and French (1993): we use 30-70 breakpoints rather than 20-80 breakpoints, we exclude firms with negative book values, we exclude financial firms, and we use market equity observed in June. Not selecting 20-80 breakpoints could be defended as such breakpoints reduce portfolio breadth and could tilt towards stocks with more exposure towards a certain factor, potentially biasing the portfolio returns upward. Firms with negative book equity value might have particularly high default risk, and the relation between default risk and leverage is different for financial firms than for other firms (Fama and French (1992)). Fama and French (1992) have also made it common practice to construct size-breakpoints based on the market capitalization of firms at the end of June. Set 3 thus only leaves four choices open: whether to impose price filters (but this seems less important now that microcaps are excluded), whether to include utilities, whether to impose industry neutrality, and using dependent or independent sorts. These four choices allow 16 combinations.

Figure 6 shows the computed non-weighted non-standard errors per factor for Set 1 (the base case), Set 2 and Set 3. For each factor, we find that the non-standard error can be heavily reduced by imposing the restrictions of Set 2, i.e., the use of NYSE breakpoints, excluding microcaps, and using value-weighting. For example, the PEAD factor has a non-standard error of above 0.10 using the original set of eleven choices, which decreases to about 0.02 (a 76% decline) for Set 2. On average, across factors, we find that non-standard errors decrease by 70% when moving from Set 1 to Set 2. Set 3 does not yield a substantial additional decline in non-standard errors for most factors. For some factors, the non-standard errors are even higher for Set 3 than for Set 2. On average, Set 3 leads to a 73% reduction compared to the base case.

When keeping in mind that imposing restrictions hurts opportunities for customization, a relatively simple recommendation to reduce non-standard errors that follows from the above analysis is to consistently use NYSE breakpoints, exclude microcaps, and employ value-weighting. Of course, in some cases an argument can be made for not following this recommendation. For instance, researchers might have a particular interest in smaller firms, or they might want to study a mechanism most applicable to illiquid stocks. Providing a clear explanation for design choices that deviate from the above recommendation in such studies appears warranted.

## 3.5 Model Selection

The prior section has shown that Sharpe ratios within factors depend on a range of construction choices and that the non-standard errors surrounding portfolio sorting can be substantial. In this section, we study the implications of non-standard errors for model selection exercises. In particular, we use the maximum squared Sharpe ratio as selection criteria for ranking asset pricing models. Additionally, we consider efficient frontier expansion, economic significance, and out-of-sample estimation, following Detzel et al. (2021).

Prior model-selection studies select factors, each based on an own set of portfolio construction choices. As these construction choices differ per factor, outcomes of model-selection studies are impacted by underlying differences in the set of construction choices. In this section we construct factors under the same set of portfolio construction choices. Thereby, we remove idiosyncratic construction effects from factors, and introduce common construction effects to evaluate them on the same basis for model selection. We then show the impact of construction choices on the outcomes of these model selection exercises.

### 3.5.1 Maximum Sharpe ratio

The ability of an asset pricing model to price assets depends on the extent to which its factors span the mean-variance efficient portfolio. When the factors of a model are mean-variance efficient, no other factor or asset can be added to improve the performance of the span of the factors. Gibbons et al. (1989) show that the gain of adding test assets to a factor model can be written as:

$$Sh^2(f,\Omega) - Sh^2(f) = \alpha'\Sigma^{-1}\alpha \tag{3.3}$$

$Sh^2(f,\Omega)$ denotes the maximum squared Sharpe ratio obtained from the factors $f$ and assets $\Omega$, and $Sh^2(f)$ for $f$. $\alpha$ is a vector of intercepts obtained from regressing the assets $\Omega$ excess return on factor returns. $\Sigma^{-1}$ is the covariance matrix of residuals from these regressions. Barillas and Shanken (2017) use the maximum squared Sharpe ratio as an indicator of model quality, since it measures how close the span of a model is to the ex-post mean-variance efficient frontier. The aim is to minimize the mispricing that an asset pricing model creates, which corresponds to minimizing the outcome of equation 3.3. Barillas and Shanken (2017) argue that $Sh^2(f,\Omega) = Sh^2(\Omega)$ when $\Omega$ consists of the entire universe of assets. In that case, minimizing the outcome of equation 3.3 corresponds to maximizing $Sh^2(f)$. Hence, model selection can be examined by comparing the maximum squared Sharpe ratio across models.[16]

The typical approach in the literature has been to compare factors using their "original" construction method, thereby comparing factors without taking differences in construction method into account. We explicitly take into account the range of possible construction methods and compare factors on an "apples-to-apples" basis. Figure 7 reports the average maximum Sharpe ratio of a factor model whereby we

[16] Detzel et al. (2021) show that when (transaction) costs are ignored, model comparison based on squared Sharpe ratios favor models with high gross performance, even when trading costs are high. Hence, Appendix 3.12 includes a model comparison analysis when considering net factor returns.

average across the possible set of construction choices. Around the average, we also plot a two standard deviation spread of the Sharpe ratio of a factor model. We separate the value-weighted portfolio returns (dashed bars) and equal-weighted portfolio returns (white bars).

The average maximum Sharpe ratio for the mean-variance optimal FF5 model, using value-weighted returns, is 1.08. Replacing operating profitability with cash profitability increases this value to 1.35. Adding the momentum factor further improves the average maximum Sharpe ratio to 1.50. The optimal Q4, BS6, and DHS factor models have an average maximum Sharpe ratio of 1.37, 1.67, and 1.71, respectively. Based on these averages, the preferred model would be the DHS factor model, with the BS6 model coming very close. When factors are equally weighted, factor models have higher maximum Sharpe ratios, on average. The highest average maximum Sharpe ratios with equal-weighted returns are also obtained by the BS6 and DHS factor models, both with values of roughly 2, and now the BS6 model has a slightly higher average value.

Differences in construction choices induce non-standard errors in factor premiums and subsequently also in the maximum Sharpe ratio of factor models. The error bars indicate that the two-standard deviation spread in the maximum Sharpe ratio can be substantial. For example, for the equally-weighted BS6 model, we find a 95% confidence interval between 1.57 and 2.54. Due to the non-standard errors, model rankings may differ across different sets of construction choices. We find that in 39.9% of all choice sets, the BS6 has the largest maximum Sharpe ratio. The DHS model has the largest maximum Sharpe ratio in 60.1% of the choice sets. These results show that even though one model can have the largest maximum Sharpe ratio in the majority of the construction choice sets, a different outcome for model selection exercises can be achieved when using other choice sets. Moreover, model rankings can especially differ when researchers make differential choices across models (for example, 80-20 breakpoints for the PEAD factor but 70-30 breakpoints for factors in another model). In section 5.4, we use a bootstrap approach to further study how often one model outperforms the others.

Table 5 reports the portfolio weights that correspond to the ex-post mean-variance efficient portfolios constructed from the candidate factor models, where we average the weights across all construction methodologies. Between brackets, we report the standard deviation of the weights, based on our set of 2048 construction methods. The standard deviation can be considered as a non-standard error in the mean-variance optimal weights due to variation in construction methodologies. Since the factors are constructed in the same way, the weights can be compared directly. We find large discrepancy in optimal weights within factor models. For example, the Fama-French 5 factor model allocates 47.6% weight towards the CMA factor, on average. However, for a researcher that randomly picks a construction choice the weight on the CMA factor varies between 27.6% and 77.6% for a two standard deviation change. HML has a small average weight of 1.5% in the 5-factor model.

Interestingly, for some construction methods the HML weight is negative (-18.1% for a two standard deviation decrease) while for others it is substantially positive (21.1% for a two standard deviation increase). Hence, in some situations, it appears that one should have a short position in HML, whereas with other construction choices a mean-variance investor should hold a long position in HML. The Q4 model aims to improve on the 5-factor model by replacing the investment factors with their ROE factor, which uses more timely information (i.e., quarterly ROE data). Compared to the Fama-French models, the Q factor seems to have more stable weights, with standard deviations between 2% and 4%. For example, the I/A factor ranges between 31.7% and 46.9%, given a two standard deviation interval. The BS6 model aims to improve on the Fama-French models by adding the monthly updated value factor, which correlates more negatively with momentum. Consequently, the UMD factor receives a larger average weight of 21.1%, with a standard deviation of 4.1% across construction methods. The monthly updated value factor receives a relatively larger weight (compared to FF-models) of 26.7%, with a standard deviation of 7.3%. On average, the PEAD factor is important in the DHS model. The model allocates on average 58.1% to the PEAD factor, with a standard deviation of 6.5%.

### 3.5.2 Efficient frontier expansion

The results from the previous subsection indicate that model performance and its underlying weights depend on construction methods. In this subsection, we aim to measure the extent to which additional factors of a model "M1" to those of model "M0" expand the efficient frontier. To this end, we implement the multi-factor version of the generalized alpha of Novy-Marx and Velikov (2016). More specifically, we run a regression of the excess returns of the ex-post mean-variance efficient portfolio constructed from the union of M1 and M0 on the returns of the mean-variance efficient portfolio using the factors from model M0:

$$MVP_{M1\cup M0,t} = \alpha + \beta MVP_{M0,t} + \epsilon_t \tag{3.4}$$

Table 6 reports the results of these spanning regressions for each pair of models, averaged over all construction methods. Typically, we find that most models expand the efficient frontier when added to other models. For example, the spanning alpha of the FF5 model augmented by other models ranges between 0.13% and 0.45% per month, with t-statistics above 4.5. We especially find that adding the BS6 factors or DHS3 factors (M1) to FF models (M0) greatly improves the efficient frontier, with alphas between 0.25% and 0.45% per month.

Across construction methods, we find large standard deviations in the estimated alphas (reported within [ ]). The Q4 model, on average, expands its efficient frontier by adding other factor models. For example, adding the FF5 model to the Q4 model expands the efficient frontier, on average, with an estimated average alpha of 0.06% per month. However, the estimated alpha has a standard deviation of 0.05%. Under some construction method, the estimated alpha may thus be considerably closer to zero. Hence, in some cases it may appear that adding one factor model to other factor models expands the efficient frontier, whereas in other cases the marginal benefit of adding a factor is small or even zero. Again, our results imply that construction methods can influence model selection exercises, as indicated by the relatively large standard deviations around the spanning alphas.

### 3.5.3 Economic significance

Next, we quantify the economic significance, in table 7, by reporting by which percentage the maximum Sharpe ratio would increase if we would add the additional factors (M1) to the base model (M0) for each pair. This exercise relates to the gain that could be realized by a mean-variance investor. In most cases, adding one model to a base model improves the Sharpe ratio of the combined model. We find that the FF5 model can be improved between 20.5% up to 82.3%, on average, by adding one of the other factor models. Adding the BS6 model to any of the FF-models could improve the maximum Sharpe ratio by between 37.1% and 82.3%, whereas adding the DHS3 model yields gains between 36.2% and 74.7%. These results indicate that the FF models are not able to span the information contained in the BS6 and DHS3 models. Adding the BS6 model to the DHS3 model yields an average improvement of 26.0%, whereas vice versa the gain is 26.9%.

The economic gain could also depend on the specific construction choice. Between parentheses, we report the standard deviation of the improvements in Sharpe ratios, across construction methods. For example, on average, the BS6 model improves the FF5 model by 82.3%, but also has a substantial standard deviation of 17.0%. This implies that there is a construction set for which the improvement is 48.3%, but also a set for which the improvement is 116.3%. The Q-factor model improves the $FF6_c$ by 6.7% on average, with a standard deviation of 5.8%. Hence, in some cases, it may appear that the economic gain is (close to) zero, thereby giving the illusion that the Q4 model is not able to improve the $FF6_c$ model. The main takeaway is that the improvement in Sharpe ratio, when adding additional factors, is not only a function of expected returns, variances, and correlations among factors, but also a function of factor construction choices.

### 3.5.4 In-sample and out-of-sample estimation

We have used full-sample estimates to calculate maximum Sharpe ratios for our model selection exercise. When factors have high average returns relative to expected returns, these factors obtain too much weight in the ex-post mean-variance tangency portfolio. The optimal mean-variance efficient weights will be overfit, even though they are noisy estimates of the true weights. Consequently, the estimates of the maximum Sharpe ratio can be biased upwards. This bias becomes larger in smaller samples, since the parameter estimates have more sampling error. Also, the bias in the estimates of the maximum Sharpe ratio is especially problematic for comparing non-nested models, such as the Q-factor model versus Fama-French models. To solve this problem, we run bootstrap simulations of in-sample (IS) and out-of-sample (OS) Sharpe ratio estimates, following Fama and French (2018). The bootstrap approach has the advantage, compared to the full-sample approach, that it is able to yield a distribution of maximum Sharpe ratio estimates and that it allows for testing how often one model outperforms the other.

The bootstrap procedure that we use is to split the 600 months into 300 adjacent pairs of months for a given set of factors constructed from construction rule $r$. For each simulation run, we draw (with replacement) a random sample of 300 pairs. We randomly assign a month from each pair to the IS sample.[17] Using this IS sample of factor returns, we compute the maximum Sharpe ratio for each model and the corresponding mean-variance optimal portfolio weights. We allocate the remaining unassigned months to the OS sample. Subsequently, we compute the out-of-sample Sharpe ratio estimate using the OS sample of factor returns and the weights estimated from the IS sample. The IS estimates are, like the full-sample estimates, subject to an upward bias. However, this is less of a problem for OS Sharpe ratios, since monthly returns are approximately serially uncorrelated. For each construction rule $r$ we run 100.000 simulation runs. For each run, we compare the maximum Sharpe ratio between models and count how many times a model has a higher maximum Sharpe ratio than an other model. By doing so, we can calculate both the in-sample and out-of-sample probability that a model is winning from other models. In addition, we can calculate this win-probability within simulation $r$ and the total win-probability averaged across all construction rules.

Table 8 shows the win-probability estimates obtained from the bootstrap simulations. Panel A shows the in-sample estimates, which should be interpreted with caution as the in-sample Sharpe ratios are upward biased and based on 300 observation months. We find that the $FF6_c$ model outperforms the Q factor model in 64.8% of the sample. Its Sharpe ratio (1.67) is slightly higher than that of the Q factor model (1.60). The BS6 model seems to outperform the other models, with pairwise win-probabilities over 50.7% and an average Sharpe ratio 0f 1.98. It is, on average, the model with the highest Sharpe ratio in 47.6% of all simulation runs. The standard deviation is 29.1%, implying large variation across construction methods. The DHS model in this simulation has an average Sharpe ratio of 1.92, making it the second-best model in this aspect. Still, this model has the largest Sharpe ratio in 47.8% of all simulation runs.

Panel B presents the out-of-sample (OS) results. We find that the DHS model outperforms all other models in 57.9% of the simulation runs in an out-of-sample setting, averaged across construction methods. The BS6 model obtains an overall win-probability of 38.4%, making it the second strongest model from an out-of-sample

[17]Note that a month might appear multiple times in the IS sample if the pair is drawn multiple times.

perspective. Both models have an almost 30% standard deviation in the overall win-probability. This implies that in many construction choices one model may appear superior to the other, and vice versa, while other models, such as the $FF6_c$ model, also have a win-probability exceeding zero. Given the high standard deviations, our conclusion is again that one should be cautious when drawing inferences from one or a few sets of construction choices.

## 3.6 Portfolio Characteristics across Construction Choices

We have shown that factor returns vary significantly across different sets of construction choices and that different construction choices can have an influence on model selection exercises. In this section, we study how variation in construction choices affects portfolio characteristics that, in turn, have an impact on portfolio performance. We consider the factor exposure, illiquidity and transaction costs of a portfolio.

Regarding factor exposure, the expected return of a well-diversified factor portfolio is directly related to the sorting characteristic (Cochrane, 2011):

$$E(R_{long} - R_{short}) = \beta(F_{long} - F_{short})$$

where $F$ stands for the factor characteristics of the long and short portfolio. We define factor exposure by creating a normalized factor score. For every variable $v$, we first compute the cross-sectional average, maximum and minimum at time $t$. Next, for every stock $i$, we compute the normalized factor score for all variables $v$ at time $t$ by subtracting the cross-sectional average from the variable score of the stock $variable_{i,v}$ and subsequently dividing by the spread between the maximum variable score in that month and the minimum variable score in that month:[18]

$$Normalized\ factor\ score_{i,v,t} = \frac{Variable_{i,v,t} - Mean_{v,t}}{Max_{v,t} - Min_{v,t}} \tag{3.5}$$

In both the long and the short side of the long-short portfolio, we aggregate the normalized factor scores to the portfolio level by using the respective weighting scheme, value- or equal-weighted. Subsequently, we compute the spread between the long and short leg of the factor portfolio to arrive at the factor exposure per factor per construction choice.

In addition to an impact on factor exposure, construction choices may impact the liquidity of a portfolio. Stocks with low liquidity, such as microcaps, may have high transaction costs and other frictions (such as relatively high bid-ask spreads), which could directly impact the returns of factor portfolios. We measure the liquidity of the portfolios by aggregating stock-level illiquidity to portfolio-level illiquidity following

[18] We calculate the normalized factor score before using exclusion criteria.

Amihud (2002). More specifically, we measure stock-level illiquidity as the average ratio of the daily absolute return to the dollar trading volume on month $t$:

$$ILLIQ_{i,t} = \frac{1}{D_{i,t}} \sum_{t=1}^{D_{i,t}} \frac{|R_{i,t,d}|}{VOLD_{i,t,d}} \tag{3.6}$$

The daily return of a stock is denoted by $R_{i,t,d}$. $VOL_{i,t,d} * P$ equals the dollar trading volume for stock $i$ on day $d$ of month $t$. $D_{i,t}$ equals the amount of trading days for stock $i$ on month $t$. A lower value of $ILLIQ_{i,t}$ implies a higher level of liquidity.

We further consider whether construction choices affect transaction costs. We estimate transaction costs at the individual stock-level using the procedure of Hasbrouck (2009). This procedure allows us to estimate effective spreads for individual stocks using their daily price series. We provide more details on this procedure in Appendix 3.12.2. To examine the impact of construction choices on factor exposure, illiquidity and transaction costs, we run fixed-effect panel regressions where we regress the constructed variables on dummy variables of each construction choice. We include factor and time fixed effects in the estimation. Table 9 shows the estimated coefficients.

Overall, most construction choices significantly impact portfolio characteristics. We find that 7, 7 and 10 out of the 11 construction choices show significant coefficients (at the 5% level) on factor exposure, portfolio illiquidity, and transaction costs, respectively. Portfolios based on 30-70 breakpoints have significantly lower factor exposures than those with 20-80 breakpoints, while they are more liquid. Furthermore, 30-70 portfolios have, on average, 6 basis points lower transaction costs than 20-80 breakpoints portfolios. Using NYSE instead of NAN breakpoints significantly lowers transaction costs by an average of 15 basis points and improves portfolio liquidity. This is sensible as NAN breakpoints allow more small firms to enter the portfolio, hence increasing transaction costs and illiquidity.

Excluding stocks with a price below 5 dollars has a significant negatively impact on factor exposures, while at the same time improving liquidity and reducing transaction costs. Including financial firms and utility firms also reduces transaction costs, albeit only with 1 basis point. Value-weighting as opposed to equal-weighting significantly reduces factor exposure. This reduction is compensated by a significantly higher liquidity profile and significantly lower transaction costs.

## 3.7 Conclusion

Within empirical asset pricing, character-based sorting is a popular way to construct factors. This paper stresses that constructing factors involves a large number of choices, leading to "degrees of freedom" for researchers. Especially since there is no consensus on construction methods, the degrees of freedom involved allows for p-hacking if the choices affect outcomes: researchers could then pick construction choices in such a way that the resulting factor meet certain statistical and performance-related hurdles, such as high Sharpe ratios.

We find that construction choices indeed impact factor returns. Using 2048 different combinations of construction choices, we show large and significant variation in Sharpe ratios based on factor returns. As such, the variation in choices for factor construction by researchers leads to substantial variation in outcomes. We calculate non-standard errors as the standard deviation of the generated Sharpe ratios and show that the non-standard errors in our setting are sizable, also in comparison with standard errors. An alternative calculation of non-standard errors that takes the popularity of choices into account reinforces this conclusion.

The variation that we document materially impacts model selection exercises when comparing models. Maximum Sharpe ratios of factor models show wide variation across construction methods and also mean-variance weights vary substantially across construction methods. By following a bootstrapping approach, we show that design choices substantially affect a model's probability of producing the highest Sharpe ratio. Our analysis indicates that factor models should not be compared against each other when their construction method differ and that it is important to check how the winning model depends on the construction choices being made.

As in Mitton (2022), who focuses on methodological variation in empirical corporate finance, we argue that the field benefits by reducing researcher latitude regarding the robustness results that are reported. Our results suggest that the most important design choice around factor construction are those concerning NYSE or NAN breakpoints, micro stocks, industry-adjusted characteristics, and value-weighting. In a specification check (Brodeur et al. (2020)), researchers could graphically show the distribution of their Sharpe ratios (or other results) if their design choices are varied among these four dimensions.

Although variation in design choices allows researchers to customize samples and tests for specific research questions, while also reducing the chance of missed discoveries, too much variation severely complicates the comparison of results across papers. Based on our analysis, we recommend the consistent use of NYSE breakpoints, exclusion of microcaps, and value-weighting. Following this guideline reduces the average non-standard error by 70%. Currently, this practice is not yet standard in the literature.

For example, when we consider the 25 publications in the Journal of Finance, Review of Financial Studies, and Journal of Financial Economics in 2021 and 2022 that employ portfolio sorting, only four of these papers use NYSE breakpoints rather than NAN breakpoints and only six papers exclude microcaps. Six of these 25 studies use equal-weighting, the use of NAN breakpoints, and micro-caps inclusion simultaneously.

Our message is that by keeping a very limited set of construction choices fixed, empirical asset pricing researchers can greatly facilitate the interpretability of their presented results, as many of the design choices that we study have more moderate effects. Stressing this more comforting message to the empirical asset pricing field is also important.

## 3.8 References

Akey, P., Robertson, A. Z., & Simutin, M. (2022). Noisy factors. *Available at SSRN 3959116*.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, *61*(1), 259–299.

Asness, C. S., Porter, R. B., & Stevens, R. L. (2000). Predicting stock returns using industry-relative firm characteristics. *Available at SSRN 213872*.

Ball, R., Gerakos, J., Linnainmaa, J. T., & Nikolaev, V. (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics*, *121*(1), 28–45.

Barillas, F., Kan, R., Robotti, C., & Shanken, J. (2020). Model comparison with sharpe ratios. *Journal of Financial and Quantitative Analysis*, *55*(6), 1840–1874.

Barillas, F., & Shanken, J. (2017). Which alpha? *The Review of Financial Studies*, *30*(4), 1316–1338.

Barillas, F., & Shanken, J. (2018). Comparing asset pricing models. *The Journal of Finance*, *73*(2), 715–754.

Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: p-hacking and publication bias in causal analysis in economics. *American Economic Review*, *110*(11), 3634–3660.

Brown, S. J., Lajbcygier, P., & Li, B. (2008). Going negative: What to do with negative book equity stocks. *Journal of Portfolio Management*, *35*(1), 95–102.

Chen, A. Y., & Zimmermann, T. (2022). Open source cross-sectional asset pricing. *Critical Finance Review, forthcoming*.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance*, *66*(4), 1047–1108.

Daniel, K., Hirshleifer, D., & Sun, L. (2020). Short-and long-horizon behavioral factors. *The Review of Financial Studies*, *33*(4), 1673–1736.

Daniel, K., Mota, L., Rottke, S., & Santos, T. (2020). The cross-section of risk and returns. *The Review of Financial Studies*, *33*(5), 1927–1979.

Detzel, A. L., Novy-Marx, R., & Velikov, M. (2021). Model selection with transaction costs. *Available at SSRN 3805379*.

Drechsler, I., & Drechsler, Q. F. (2014). *The shorting premium and asset pricing anomalies* (Tech. Rep.). National Bureau of Economic Research.

Ehsani, S., Harvey, C. R., & Li, F. (2021). Is sector-neutrality in factor investing a mistake? *Available at SSRN 3930228*.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2008). Dissecting anomalies. *The Journal of Finance*, *63*(4), 1653–1678.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, *128*(2), 234–252.

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, 1121–1152.

Green, J., Hand, J. R., & Zhang, X. F. (2013). The supraview of return predictive signals. *Review of Accounting Studies*, *18*, 692–730.

Harvey, C. R. (2017). Presidential address: The scientific outlook in financial economics. *The Journal of Finance*, *72*, 1399–1440.

Harvey, C. R., & Liu, Y. (2019). A census of the factor zoo. *Available at SSRN 3341728*.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68.

Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *The Journal of Finance*, *64*(3), 1445–1477.

Hasler, M. (2022a). Is the value premium smaller than we thought? *Critical Finance Review, forthcoming*.

Hasler, M. (2022b). Looking under the hood of data-mining. *Available at SSRN*.

Hirshleifer, D., & Jiang, D. (2010). A financing-based misvaluation factor and the cros-section of expected returns. *Review of Financial Studies*, *23*(1), 3401–3436.

Hou, K., Mo, H., Xue, C., & Zhang, L. (2019). Which factors? *Review of Finance*, *23*(1), 1–35.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, *28*(3), 650–705.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, *33*(5), 2019–2133.

Jensen, T. I., Kelly, B. T., & Pedersen, L. H. (2022). Is there a replication crisis in finance? *Journal of Finance, forthcoming*.

Kessler, S., Scherer, B., & Harries, J. P. (2020). Value by design? *Journal of Portfolio Management*, *46*(2), 1–19.

Lee, C., & Swaminathan, B. (2000). Price momentum and trading volume. *Journal of Finance*, *55*(1), 2017–2069.

Linnainmaa, J. T., & Roberts, M. R. (2018). The history of the cross-section of stock returns. *The Review of Financial Studies*, *31*(7), 2606–2649.

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, *71*(1), 5–32.

Menkveld, A. J., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., ... Weitzel, U. (2023). Non-standard errors. *The Journal of Finance, forthcoming*.

Mitton, T. (2022). Methodological variation in empirical corporate finance. *The Review of Financial Studies*, *35*, 527–575.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, *108*(1), 1–28.

Novy-Marx, R., & Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, *29*(1), 104–147.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, *39*(4), 1127–1139.

Soebhag, A., Van Vliet, B., & Verwijmeren, P. (2022). Mind your sorts. *Available at SSRN*.

Stambaugh, R. F., & Yuan, Y. (2017). Mispricing factors. *The Review of Financial Studies*, *30*(4), 1270–1315.

Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics*, *99*(1), 1–10.

Wahal, S., & Yavuz, M. D. (2013). Style investing, comovement and return predictability. *Journal of Financial Economics*, *107*(1), 136–154.

Walter, D., Weber, R., & Weiss, P. (2022). Non-standard errors in portfolio sorts. *Available at SSRN*.

Yan, X. S., & Zheng, L. (2017). Fundamental analysis and the cross-section of stock returns: A data-mining approach. *The Review of Financial Studies*, *30*(4), 1382–1423.

## 3.9 Figures & Tables

Figure 1: **Construction choices and Sharpe ratios of the HML factor.** This figure plots annualized gross Sharpe ratios (y-axis) for long-short factor returns, where a factor is constructed by using 2048 different factor construction methods. The x-axis shows the 2048 different versions of the value factor ordered from low Sharpe ratios to high. The red dot shows the median Sharpe ratio for the HML factor in our sample. The blue dot shows the Sharpe ratio using the construction choices mentioned in the original study. The sample runs from January 1972 until December 2021.

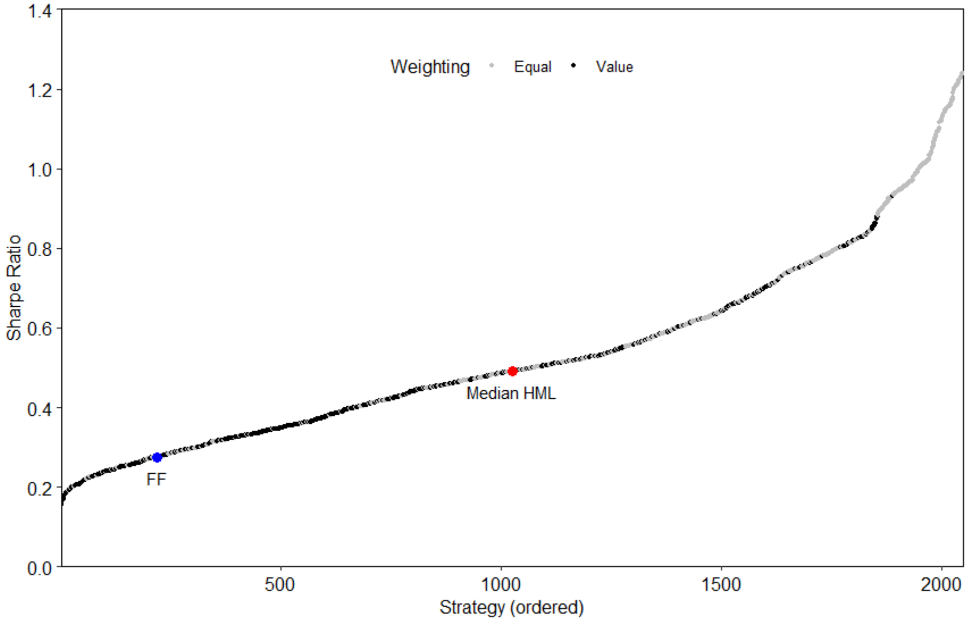
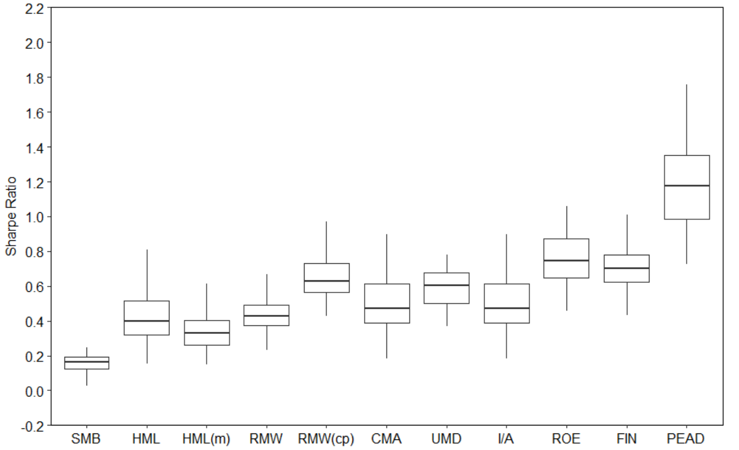
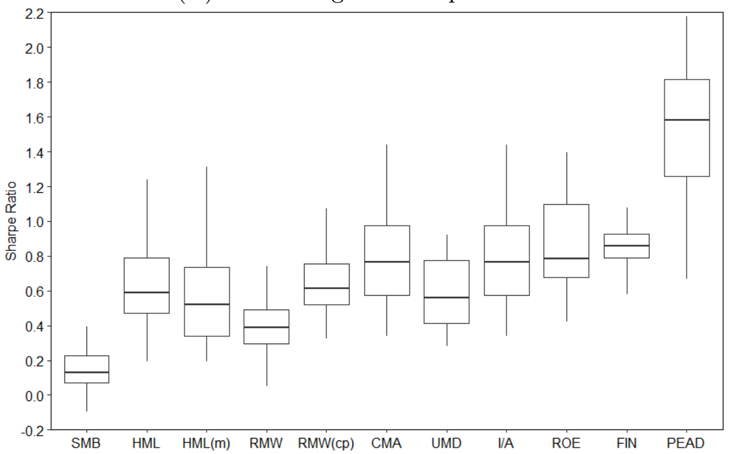


Figure 2: **Sharpe ratio variation within factors.** This figure plots the distribution of annualized value-weighted (subfigure A) and equal-weighted (subfigure A) gross Sharpe ratios for long-short factor returns, where a factor is constructed 2048 times by using the 2048 different factor construction methods. The black solid line within the box plot shows the median Sharpe ratio. The upper (lower) bound shows the 75th (25th) percentile. The factors and their definitions are from Table 1. The sample runs from January 1972 until December 2021.




(A) Value-Weighted Sharpe Ratios




(B) Equal-Weighted Sharpe Ratios

Figure 3: **Construction choices and Sharpe ratios.** This figure shows the impact of construction choices on the gross Sharpe ratio averaged over factors. Sharpe ratios are annualized. We consider eleven choices. “30-70” refers to the use of the 30th and 70th percentile as breakpoints in the sorting procedure (“Yes”) or the use of the 20th and 80th percentile (“No”). NYSE indicates whether the NYSE cross-section is used to construct breakpoints (“Yes”) or the full NYSE-AMEX-Nasdaq cross-section (“No”). “BE” indicates whether stocks with negative book equity are excluded (“Yes”) or included (“No”). “Micro” indicates whether we include stocks with the smallest 20% market capitalization (“Yes”) or not (“No”). “PRC” indicates whether stocks with a price below 5 dollar are excluded (“Yes”) or included (“No”). “Utilities” means that companies in the utility sector are included (“Yes”) or excluded (“No”). “Financial” means that companies in the finance sector are included (“Yes”) or excluded (“No”). “Ind_Neutral” means that portfolio sorts are constructed using industry-adjusted characteristics (“Yes”) or the standard characteristics (“No”). “VW” indicates whether factors are calculated by using value-weighting (“Yes”) or equal-weighting (“No”). “Independent” refers to the use of independent sorting (“Yes”) or dependent sorting (“No”). “Recent” indicates that we use the one-month lagged market capitalization (“Yes”) or the market capitalization of June (“No”). Monthly factor returns are constructed using data from January 1972 to December 2021.

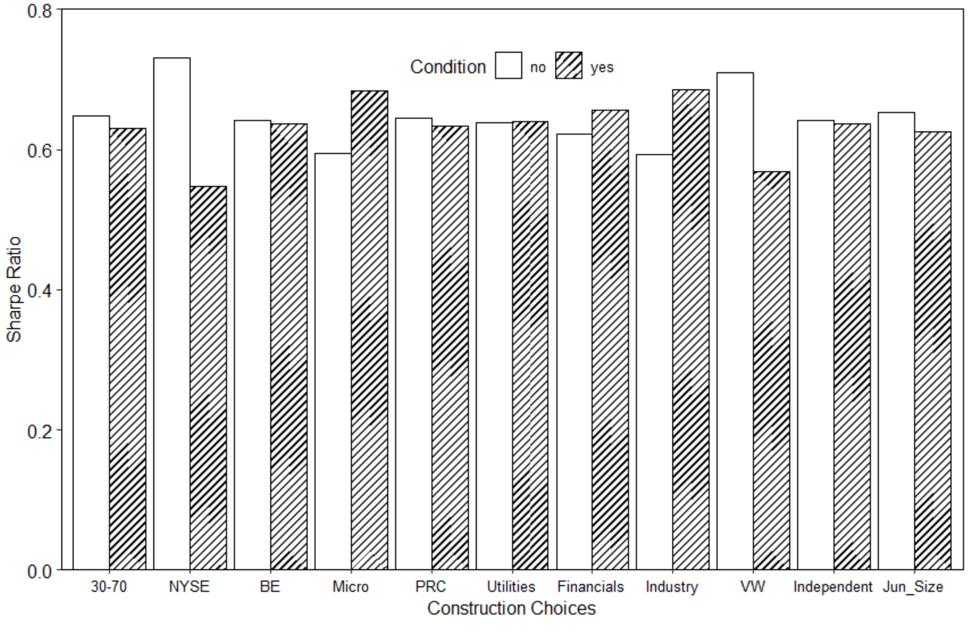


Figure 4: **Non-standard errors and standard errors.** This figure plots the non-standard error (white) and standard error (dashed bar) for each factor. The non-standard error is defined as the cross-sectional standard deviation of Sharpe ratios, where the cross-section consist of all 2048 versions of a factor. The standard error is the standard deviation of the Sharpe ratio obtained from block-bootstrapping a factor, averaged over the construction choices. We block-bootstrap each series 10.000 times. The error line on the dashed bar indicates the minimum and maximum standard error within a factor. Monthly factor returns are constructed using data spanning January 1972 and December 2021.

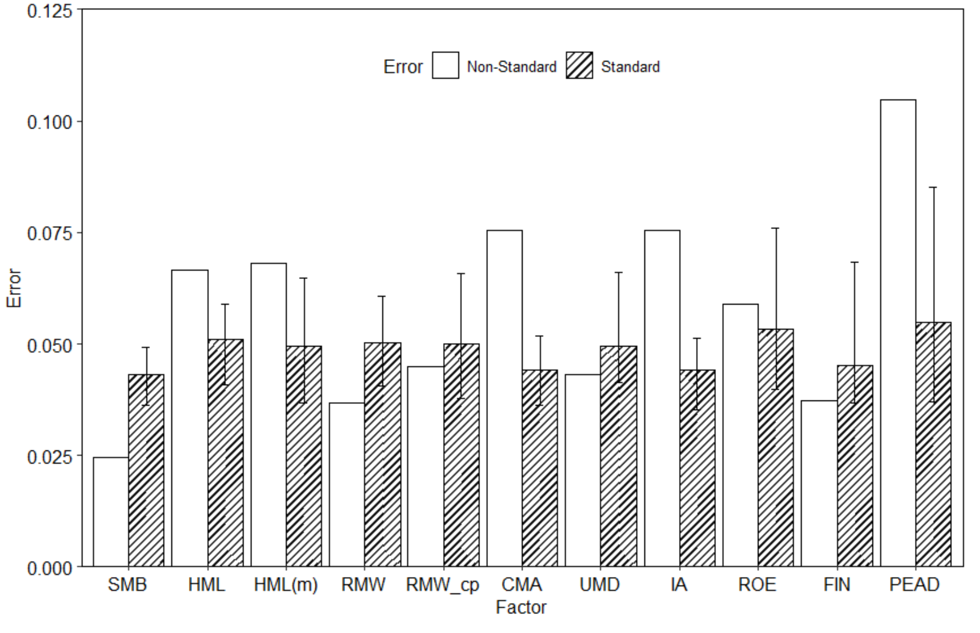


Figure 5: **Weighted non-standard errors and standard errors.** This figure plots the non-standard error (white) and standard error (dashed bar) for each factor. The non-standard error is defined as the cross-sectional weighted standard deviation of Sharpe ratios, where the cross-section consist of all 2048 versions of a factor. The standard error is the weighted standard deviation of the Sharpe ratio obtained from block-bootstrapping a factor, averaged over the construction choices. We block-bootstrap each series 10.000 times. We weight the errors by the survey-implied probabilities. The error line on the dashed bar indicates the minimum and maximum standard error within a factor. Monthly factor returns are constructed using data spanning January 1972 and December 2021.

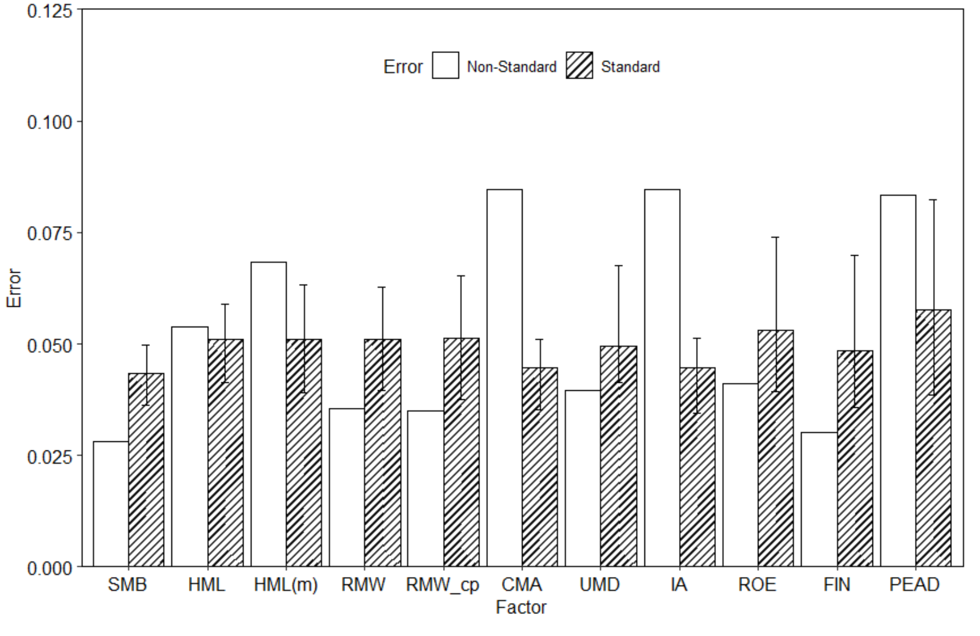


Figure 6: **Reducing non-standard errors.** The figure shows the non-standard errors using three sets of research design choices. Set (1) includes all eleven construction choices. Set (2) imposes NYSE breakpoints, excludes micro-caps, and imposes value-weighting. Set (3) extends on set (2) by further imposing the use of 30-70 breakpoints, the exclusion of firms with a negative book value and financial firms, and by imposing the measurement of size in June.

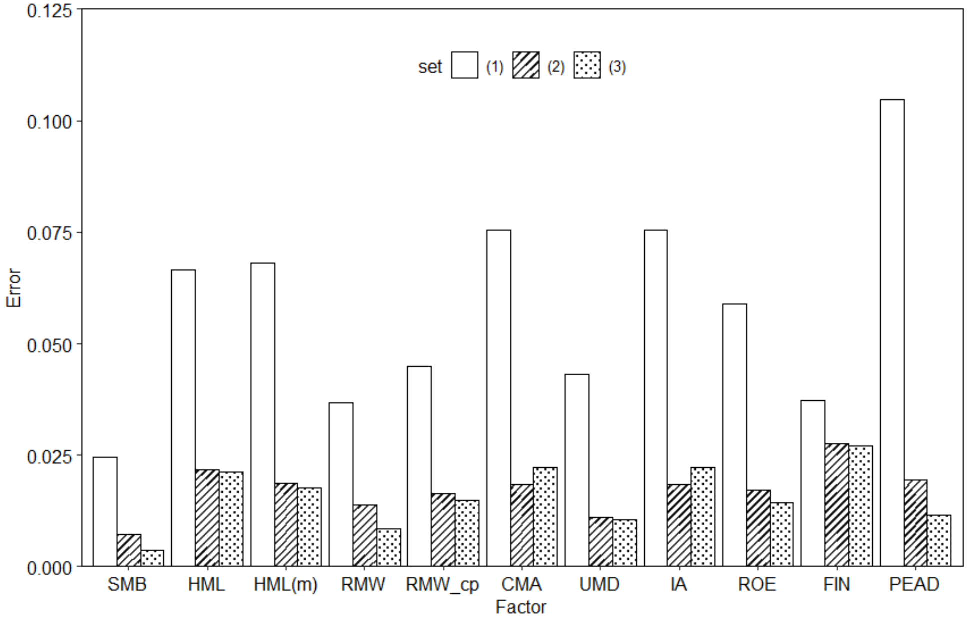


Figure 7: **Selecting factor models.** This figure shows the maximum gross Sharpe ratio (annualized) from the factors from the factor models listed on the horizontal axis. The white bar shows the maximum Sharpe ratio obtained by using equal weighted factor returns. The dashed bar shows the maximum Sharpe ratio using value weighted factor returns. The error plot shows the variation (95% confidence interval) in the maximum Sharpe ratios for a given factor model, across construction choices. The data runs from January 1972 until December 2021.

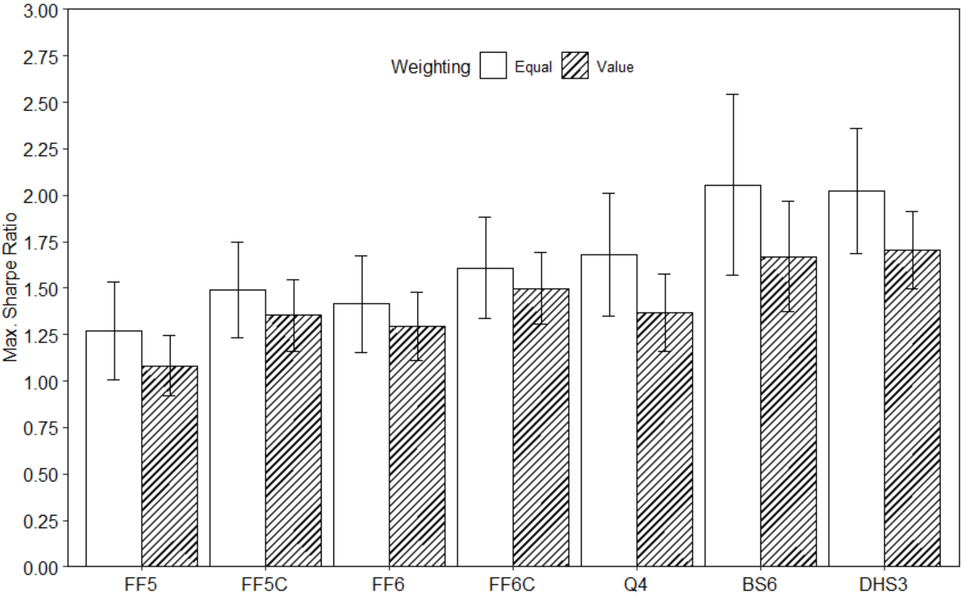


Table 1: **Factor models.** This table lists the non-market factors used by asset pricing models, indicated by a ✓. It also lists properties of the factor construction methodology: the sorting characteristic, the breakpoints (BP), the rebalancing frequency (Rebalancing), and the sorting method (Construction). In each model, factor returns are defined as the equal-weighted average of the returns on the portfolios with high (or low) values of the primary sorting characteristic minus the equal-weighted average of the portfolios with low (or high) values. SMB returns are given by the simple average of the returns on all portfolios with low size minus the average of the returns on all portfolios with large size in three independent 2x3 sorts of stocks on size and each of the following characteristics: book-to-market, growth in book assets, and operating profitability. ME returns are given by the simple average of the returns on all portfolios with low size minus those with large size in 2x3x3 sorts on size, growth in book assets, and return on equity. FF5 (FF6) denote the Fama and French (2015) five-factor model (augmented with UMD). $FF5_c$ and $FF6_c$ denote versions of the FF5 and FF5M, respectively, that use cash-based operating profitability instead of accruals operating profitability, based on Fama and French (2018). Q4 denotes the Hou et al. (2015) four-factor q-model. BS6 denotes the Barillas and Shanken (2018) six-factor model. DHS3 denotes the Daniel, Hirshleifer, and Sun (2020) three-factor model.

| | | | | | Factor models | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Factor | Sorting characteristic | BP | Rebalancing | Construction | FF5 | FF6 | $FF5_c$ | $FF6_c$ | Q4 | BS6 | DHS3 |
| SMB | Market capitalization | 50-50 | Annual | 2x3 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| HML | Book-to-market | 70-30 | Annual | 2x3 | ✓ | ✓ | ✓ | ✓ | | | |
| HML(m) | Book-to-market | 70-30 | Monthly | 2x3 | | | | | | ✓ | |
| RMW | Accruals operating profitability | 70-30 | Annual | 2x3 | ✓ | ✓ | | | | | |
| RMW(cp) | Cash operating profitability | 70-30 | Annual | 2x3 | | | ✓ | ✓ | | | |
| CMA | Growth in book assets | 70-30 | Annual | 2x3 | ✓ | ✓ | ✓ | ✓ | | | |
| UMD | $R_{t-12,t-2}$ | 70-30 | Monthly | 2x3 | | ✓ | | ✓ | | ✓ | |
| I/A | Growth in book assets | 70-30 | Monthly | 2x3x3 | | | | | ✓ | ✓ | |
| ROE | Quarterly returns-on-equity | 70-30 | Monthly | 2x3x3 | | | | | ✓ | ✓ | |
| FIN | Net and composite share issuance | 80-20 | Annual | 2x3 | | | | | | | ✓ |
| PEAD | 4-day CAR earnings announcements | 80-20 | Monthly | 2x3 | | | | | | | ✓ |

Table 2: **Variation in empirical finance.** This table shows the results from surveying the main methodological choices that have been made in the empirical asset pricing literature. We report the proportions of the choice (1) or (2) occuring in 323 empirical articles in the top finance journals between 1965 and 2018, based on the list of papers that Campbell Harvey and Yan Liu constructed for their census of the factor zoo (Harvey and Liu (2019)).

| | **Options** | | **Proportion** | |
|---|---|---|---|---|
| | (1) | (2) | (1) | (2) |
| Choice 1 | Use 30-70 BP | Use 20-80 BP | 49.3% | 50.7% |
| Choice 2 | Use NYSE BP | Use NAN BP | 41.5% | 58.5% |
| Choice 3 | Exclude $BE < 0$ | Include $BE < 0$ | 22.0% | 78.0% |
| Choice 4 | Include Microcaps | Exclude Microcaps | 88.2% | 11.8% |
| Choice 5 | Impose price filter | No price filter | 18.3% | 81.7% |
| Choice 6 | Include utilities | Exclude utilities | 90.1% | 9.9% |
| Choice 7 | Include financials | Exclude financials | 71.2% | 28.8% |
| Choice 8 | Industry Neutrality | Unhedged | 11.5% | 88.5% |
| Choice 9 | Value-Weighted | Equal-Weighted | 58.5% | 41.5% |
| Choice 10 | Independent | Dependent | 71.8% | 28.2% |
| Choice 11 | June Size | Recent Size | 67.4% | 32.6% |

Table 3: **Correlation of choices.** This table reports the correlation matrix between the eleven methodological choices in our sample.

| Choice | 30-70 | NYSE | BE | Micro | PRC | Utilities | Financials | Industry | VW | Independent | Jun_Size |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 30-70 | 1 | | | | | | | | | | |
| NYSE | 0.12 | 1 | | | | | | | | | |
| BE | -0.05 | 0.32 | 1 | | | | | | | | |
| Micro | 0.02 | 0 | 0.02 | 1 | | | | | | | |
| PRC | 0.07 | -0.14 | -0.19 | -0.42 | 1 | | | | | | |
| Utilities | -0.08 | 0.18 | 0.15 | -0.1 | 0.03 | 1 | | | | | |
| Financials | -0.17 | -0.30 | -0.53 | -0.14 | 0.25 | 0.32 | 1 | | | | |
| Industry | 0.06 | -0.18 | -0.12 | -0.19 | 0.21 | -0.05 | 0.12 | 1 | | | |
| VW | 0.12 | 0.19 | 0.24 | 0.22 | -0.15 | -0.17 | -0.3 | -0.12 | 1 | | |
| Independent | 0.30 | 0.14 | 0.26 | -0.02 | -0.06 | 0.11 | -0.31 | 0.03 | 0.14 | 1 | |
| Jun_Size | 0.28 | 0.27 | 0.38 | 0.19 | -0.23 | 0.07 | -0.43 | -0.09 | 0.31 | 0.54 | 1 |

Table 4: **Summary statistics.** This table reports the annualized average return (in %) and Sharpe ratio of the factors listed in Table 1, gross of transaction costs. We report these statistics for both the value-weighted and equal-weighted models. The data runs from January 1972 until December 2021.

| | **Value-Weighted** | | **Equal-Weighted** | |
|---|---|---|---|---|
| | Mean | Sharpe | Mean | Sharpe |
| SMB | 1.91 | 0.18 | 0.86 | 0.11 |
| HML | 3.37 | 0.36 | 4.85 | 0.53 |
| HML(m) | 3.37 | 0.30 | 4.66 | 0.42 |
| RMW | 3.84 | 0.47 | 4.13 | 0.50 |
| RMW(cp) | 4.66 | 0.70 | 5.14 | 0.77 |
| CMA | 3.09 | 0.46 | 3.96 | 0.64 |
| UMD | 8.00 | 0.59 | 9.08 | 0.70 |
| IA | 3.09 | 0.46 | 3.96 | 0.64 |
| ROE | 7.43 | 0.80 | 9.14 | 1.03 |
| FIN | 7.51 | 0.72 | 8.90 | 0.89 |
| PEAD | 5.96 | 1.10 | 7.01 | 1.55 |

Table 5: **Mean-variance efficient portfolio weights.** This table shows the optimal weights that a mean-variance efficient investor would allocate to factors within a factor model, averaged over our set of possible construction methodologies. Within brackets, we show the standard deviation of the optimal weights that occur within our set of possible construction methods. The table shows the weights using factor returns gross of transaction costs. The sample period is from January 1972 to December 2021.

| | Mkt | SMB | HML | RMW | CMA | UMD | $RMW_{cp}$ | IA | ROE | $HML_d$ | FIN | PEAD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FF5 | 20.2 | 5.6 | 1.5 | 25.1 | 47.6 | | | | | | | |
| | (3.9) | (3.0) | (9.8) | (6.4) | (10.0) | | | | | | | |
| FF6 | 17.9 | 6.2 | 9.7 | 20.6 | 29.3 | 16.3 | | | | | | |
| | (3.6) | (3.2) | (8.8) | (6.2) | (10.2) | (3.8) | | | | | | |
| FF5C | 18.1 | 9.4 | -0.5 | | 33.3 | | 39.7 | | | | | |
| | (3.1) | (2.8) | (9.3) | | (10.7) | | (7.4) | | | | | |
| FF6C | 16.8 | 9.1 | 6.4 | | 21.5 | 13.0 | 33.2 | | | | | |
| | (3.0) | (2.7) | (8.7) | | (9.9) | (3.2) | (6.6) | | | | | |
| Q4 | 17.3 | 11.2 | | | | | | 39.3 | 32.1 | | | |
| | (2.5) | (2.0) | | | | | | (3.8) | (3.8) | | | |
| BS6 | 14.4 | 7.2 | | | | 21.1 | | 9.6 | 21.0 | 26.7 | | |
| | (3.3) | (3.0) | | | | (4.1) | | (7.0) | (4.5) | (7.3) | | |
| DHS3 | 17.7 | | | | | | | | | | 24.1 | 58.1 |
| | (2.8) | | | | | | | | | | (5.0) | (6.5) |

Table 6: **Frontier expansion.** This table reports the intercepts obtained from the regression $MVE_{M1UM0,t} = \alpha + \beta MVE_{M0,t} + \epsilon_t$. M0 is the "base" model, which is augmented to model $M1UM0$ by adding the factors of $M1$ to $M0$. $MVE_{M1UM0,t}$ is the corresponding mean-variance efficient portfolio obtained from the union of factors of $M1$ and $M0$. $MVE_{M0,t}$ is the mean-variance efficient portfolio of the factors from model $M0$. The t-statistics, reported within parentheses, are heteroskedasticity robust. Within brackets, we report the cross-sectional standard deviation of alpha. The table reports the results using gross returns. The data runs from January 1972 until December 2021.

| | **Union Model (M1)** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Base Model (M0)** | FF5 | $FF5_c$ | FF6 | $FF6_c$ | Q4 | BS6 | DHS3 |
| FF5 | 0.00 | 0.13 | 0.14 | 0.21 | 0.27 | 0.45 | 0.36 |
| | (0.00) | (4.89) | (4.86) | (6.79) | (7.48) | (11.87) | (10.59) |
| | [0.00] | [0.04] | [0.05] | [0.06] | [0.11] | [0.15] | [0.09] |
| $FF5_C$ | 0.03 | 0.00 | 0.13 | 0.11 | 0.13 | 0.35 | 0.31 |
| | (2.13) | (0.00) | (4.93) | (4.39) | (5.28) | (10.50) | (9.73) |
| | [0.03] | [0.00] | [0.05] | [0.04] | [0.07] | [0.13] | [0.09] |
| FF6 | 0.00 | 0.09 | 0.00 | 0.09 | 0.16 | 0.36 | 0.29 |
| | (0.00) | (4.35) | (0.00) | (4.35) | (5.14) | (8.39) | (9.31) |
| | [0.00] | [0.03] | [0.00] | [0.03] | [0.12] | [0.16] | [0.09] |
| $FF6_c$ | 0.02 | 0.00 | 0.02 | 0.00 | 0.06 | 0.27 | 0.25 |
| | (1.82) | (0.00) | (1.82) | (0.00) | (3.37) | (7.44) | (8.81) |
| | [0.02] | [0.00] | [0.02] | [0.00] | [0.06] | [0.14] | [0.09] |
| Q4 | 0.06 | 0.03 | 0.09 | 0.08 | 0.00 | 0.18 | 0.26 |
| | (2.89) | (2.33) | (3.72) | (3.75) | (0.00) | (6.33) | (8.16) |
| | [0.05] | [0.03] | [0.06] | [0.04] | [0.00] | [0.08] | [0.07] |
| BS6 | 0.16 | 0.14 | 0.16 | 0.14 | 0.00 | 0.00 | 0.22 |
| | (5.09) | (5.07) | (5.09) | (5.07) | (0.00) | (0.00) | (8.10) |
| | [0.11] | [0.08] | [0.11] | [0.08] | [0.00] | [0.00] | [0.06] |
| DHS | 0.09 | 0.11 | 0.10 | 0.12 | 0.11 | 0.20 | 0.00 |
| | (4.23) | (5.06) | (4.31) | (5.10) | (4.79) | (7.41) | (0.00) |
| | [0.08] | [0.07] | [0.08] | [0.07] | [0.07] | [0.11] | [0.00] |

Table 7: **Economic significance.** This table reports the increase in the maximum Sharpe ratio of the augmented model $M1UM0, t$ relative to the base model $M0$, to quantify the economic significance: $\Delta\%Sh(M0, M1) = Sh(M0, M1)/Sh(M0) - 1$. The table reports the results using gross returns. The standard deviation of the increase in Sharpe ratio, across construction methods, is reported in parentheses. The data runs from January 1972 until December 2021.

| | **Union Model (M1)** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Base Model (M0)** | FF5 | $FF5_c$ | FF6 | $FF6_c$ | Q4 | BS6 | DHS3 |
| FF5 | | 20.5 | 21.8 | 36.8 | 36.6 | 82.3 | 74.7 |
| | | (9.0) | (8.4) | (12.2) | (13.1) | (17.0) | (15.5) |
| $FF5_c$ | 4.0 | | 18.0 | 15.1 | 16.5 | 57.1 | 55.6 |
| | (3.8) | | (7.0) | (6.1) | (7.8) | (16.9) | (13.5) |
| FF6 | 0.0 | 12.2 | | 12.2 | 16.0 | 50.9 | 45.0 |
| | (0.0) | (4.9) | | (4.9) | (10.6) | (21.6) | (10.1) |
| $FF6_c$ | 2.5 | 0.0 | 2.5 | | 6.7 | 37.1 | 36.2 |
| | (2.4) | (0.0) | (2.4) | | (5.8) | (19.6) | (9.2) |
| Q4 | 5.4 | 4.1 | 8.7 | 9.8 | | 22.0 | 37.3 |
| | (4.1) | (3.8) | (4.2) | (6.5) | | (10.5) | (8.5) |
| BS6 | 15.9 | 15.3 | 15.9 | 15.3 | 0.0 | | 26.9 |
| | (9.7) | (7.8) | (9.7) | (7.8) | (0.0) | | (6.2) |
| DHS3 | 9.5 | 12.8 | 10.7 | 13.8 | 11.4 | 26.0 | |
| | (7.7) | (7.6) | (8.7) | (8.4) | (7.1) | (16.2) | |

Table 8: **In-sample and out-of-sample Sharpe ratios.** This table reports the percentage of bootstrap simulations where the maximum Sharpe ratio of the model in the row exceeds that of the model in the column, averaged across construction methodologies. We use the factor models listed in Table 1. "SR" reports the maximum Sharpe ratio of the row model, averaged across construction methodologies. $\sigma(SR)$ reports the standard deviation of the maximum Sharpe ratio of the row model. "Best" reports the estimated probability that the row model produces the highest squared Sharpe ratio among all models in the run, averaged over construction methods. $\sigma(Best)$ reports the corresponding standard deviation. Panel A presents the in-sample estimates and Panel B shows the out-of-sample estimates using gross returns. The estimates are based on 100.000 in-sample and out-of-sample simulation runs. Each simulation run splits the 600 sample months, running from January 1972 until December 2021, into 300 adjacent pair-months. The run randomly draws a sample of pairs (with replacement). The in-sample simulation randomly draws one month from each pair within a run. The remaining months form the out-of-sample. The in-sample observations are used to calculate in-sample Sharpe ratios and portfolio weights. The in-sample portfolio weights are applied to the out-of-sample returns to produce an out-of-sample Sharpe ratio estimate.

**Panel A: In-sample estimates**

| | FF5 | FF6 | $FF5_c$ | $FF6_c$ | Q4 | BS6 | DHS | Best | $\sigma(Best)$ | SR | $\sigma(SR)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| FF5 | 0.0 | 1.9 | 0.0 | 0.1 | 2.2 | 0.1 | 2.8 | 0.00 | 0.00 | 1.27 | 0.23 |
| FF6 | 98.1 | 0.0 | 30.3 | 0.0 | 22.6 | 2.3 | 8.3 | 0.00 | 0.00 | 1.45 | 0.23 |
| $FF5_c$ | 100.0 | 69.7 | 0.0 | 4.4 | 38.8 | 0.9 | 10.1 | 0.04 | 0.12 | 1.54 | 0.23 |
| $FF6_c$ | 99.9 | 100.0 | 95.6 | 0.0 | 64.8 | 10.6 | 20.7 | 4.57 | 7.34 | 1.67 | 0.24 |
| Q4 | 97.8 | 77.4 | 61.2 | 35.2 | 0.0 | 0.0 | 14.0 | 0.00 | 0.00 | 1.60 | 0.32 |
| BS6 | 99.9 | 97.7 | 99.1 | 89.4 | 100.0 | 0.0 | 50.7 | 47.61 | 29.06 | 1.98 | 0.44 |
| DHS | 97.2 | 91.7 | 89.9 | 79.3 | 86.0 | 49.3 | 0.0 | 47.77 | 28.55 | 1.92 | 0.32 |

**Panel B: Out-of-sample estimates**

| | FF5 | FF6 | $FF5_c$ | $FF6_c$ | Q4 | BS6 | DHS | Best | $\sigma(Best)$ | SR | $\sigma(SR)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| FF5 | 0.0 | 3.0 | 6.5 | 1.9 | 1.3 | 0.3 | 2.1 | 0.01 | 0.04 | 1.12 | 0.25 |
| FF6 | 97.0 | 0.0 | 38.8 | 8.8 | 18.7 | 3.8 | 6.2 | 0.40 | 0.80 | 1.30 | 0.24 |
| $FF5_c$ | 93.5 | 61.2 | 0.0 | 5.4 | 24.6 | 1.2 | 5.5 | 0.02 | 0.06 | 1.35 | 0.25 |
| $FF6_c$ | 98.1 | 91.2 | 94.6 | 0.0 | 51.1 | 11.0 | 12.9 | 2.77 | 4.62 | 1.49 | 0.25 |
| Q4 | 98.7 | 81.3 | 75.4 | 48.9 | 0.0 | 5.1 | 12.3 | 0.51 | 0.91 | 1.49 | 0.33 |
| BS6 | 99.7 | 96.2 | 98.8 | 89.0 | 94.9 | 0.0 | 40.6 | 38.44 | 29.68 | 1.80 | 0.45 |
| DHS | 97.9 | 93.8 | 94.5 | 87.1 | 87.7 | 59.4 | 0.0 | 57.85 | 29.26 | 1.84 | 0.33 |

Table 9: **Portfolio characteristics.** This table shows the estimated coefficients obtained from fixed effect regressions about the relation between eleven construction choices and ex-ante long-short normalized factor exposure, portfolio illiquidity, and transaction costs. The construction choice definitions are the same as in Figure 3. The normalized factor exposures are calculated for each firm on a monthly basis and aggregated to a portfolio level. The normalized firm factor exposure is calculated with: $(Variable - Mean)/(Max - Min)$. Illiquidity is calculated following Amihud (2002) and transaction costs following Hasbrouck (2009). Monthly characteristics are constructed using data from January 1972 to December 2021. Factor fixed effects and time fixed effects are included. Observations are weighted by factor. Double-clustered (by factor and date) adjusted t-statistics are reported between parentheses (Thompson, 2011). Asterisks are used to indicate significance at a 10% (*), 5% (**), or 1% (***) level.

| | 30-70 | NYSE | BE | Micro | PRC | Utilities | Financials | Industry | VW | Independent | Jun_Size |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Factor- | -1.13*** | -0.16 | -1.15*** | -0.63*** | -1.47*** | -0.07 | -0.17** | -0.29*** | -0.39*** | 0.01 | 0.04 |
| Exposure | (-4.33) | (-0.99) | (-4.67) | (-4.58) | (-4.78) | (-1.76) | (-2.71) | (-3.81) | (-4.21) | (1.02) | (1.36) |
| Illiquidity | -0.13** | -0.89*** | -0.04** | 1.99*** | -1.95*** | -0.08*** | 0.04* | -0.01 | -1.69*** | -0.04 | -0.00 |
| | (-2.46) | (-8.91) | (-2.46) | (5.44) | (-5.43) | (-7.97) | (2.06) | (-0.53) | (-5.74) | (-0.94) | (-0.03) |
| Transaction- | -0.06*** | -0.15*** | -0.00 | 0.11*** | -0.10*** | -0.01*** | -0.01*** | 0.05*** | -0.15*** | -0.00*** | -0.02*** |
| Cost | (-5.36) | (-6.85) | (-1.08) | (7.48) | (-5.63) | (-9.29) | (-11.89) | (6.44) | (-6.85) | (-4.05) | (-5.83) |

## 3.10 Sorting Variables

This Appendix explains the sorting variables in more detail.

**Market**: Market is the return on the CRSP value weighted stock market index in excess of the risk-free rate.
**Market capitalization**: market capitalization is the price (CRSP item PRC) times shares outstanding (CRSP item SHROUT). Market capitalization is used to construct the size factor (SMB).
**Book-to-market ratio**: Book equity in the sort for June of year $t$ is defined as the total assets for the previous fiscal year-end in calendar year $t-1$, minus liabilities, plus deferred taxes and investment tax credit, minus preferred stock liquidating value if available or redemption value if available, or carrying value. The carrying value is adjusted for net share issuance from the fiscal year-end to the end of December of $t-1$. Market capitalization is price times shares outstanding at the end of December of $t-1$, from CRSP. The book-to-market ratio is used to construct the value factor (HML). The monthly updated book-to-market ratio is used to construct the monthly value factor (HML(m)).
**Growth in book assets:** Growth in book assets, in year $t$, is defined as the change in total assets from the fiscal year ending in $t-2$ to the fiscal year ending in $t-1$ divided by total assets at $t-2$. This signal is used to construct the CMA and IA factor. The subtle difference is that for CMA, we filter observations with negative annual book equity, whereas for the IA factor, it is the quarterly book equity.
**Operating Profitability:** Operating Profitability in the sort for June of year $t$ is measured with accounting data for the fiscal year ending in year $t-1$ and is revenues minus cost of goods sold, minus selling, general, and administrative expenses, minus interest expense, minus research and development expenses, all divided by book equity. This signal is used to construct the RMW factor.
**Cash Profitability:** Cash profitability is operating profitability minus accruals for the fiscal year ending in $t-1$. Accruals are the change in accounts receivable from $t-2$ to $t-1$, plus the change in prepaid expenses, minus the change in accounts payable, inventory, deferred revenue, and accrued expenses (Ball et al., 2016). This signal is used to construct the cash-based RMW factor.
**Momentum:** Momentum is the cumulative return between month $t-12$ and $t-2$, which is used to construct the UMD factor.
**Quarterly Return-on-equity:** This is the income before extraordinary items (Compustat quarterly item IBQ) divided by 1-quarter-lagged book equity. Earnings data are used in the months immediately after the most recent public quarterly earnings announcement dates (Compustat item RDQ). In addition, we require the end of the fiscal quarter that corresponds to its most recently announced quarterly earnings to be within 6 months prior to the portfolio formation, to exclude stale earnings. We use this signal to construct the ROE factor.
**Composite share issuance:** The composite share issuance is the firm's 5-year growth in market equity, minus the 5-year equity return, in logs. We use this signal, together with net share issuance, to construct the financing (FIN) factor.

**Net share issuance:** this signal is similar to the composite share issuance, except that we use a 1-year horizon and exclude cash dividends.
**Cumulative abnormal returns earnings announcement:** we compute the cumulative abnormal returns around earnings announcements as the 4-day cumulative abnormal return from day $t-2$ to $t+1$ around the latest quarterly earnings announcement date (Compustat item RDQ):

$$CAR_i = \sum_{d=-2}^{d+1} (R_{i,d} - Rm, d) \tag{3.7}$$

where $R_{i,d}$ denotes the stock return on day $d$ and $R_{m,d}$ denotes the market return. We use the cumulative abnormal return in the months immediately following the quarterly earnings announcement date, but within 6 months from the fiscal quarter end (to exclude stale earnings). We require the earnings announcement date to be after the corresponding fiscal quarter end. In addition, we require valid daily returns on at least two of the trading days in the CAR window. We also require the Compustat earnings date (RDQ) to be at least two trading days prior to the month end. We use the most recent CAR to construct the PEAD factor.

## 3.11 The 2x3x3 Sorting Procedure

The results in the main body of the paper focus on 2x3 sorts, as in Fama and French (1993). In this appendix we consider 2x3x3 sorts. For example, the IA and ROE factors from Hou et al. (2015) are constructed using the 2x3x3 independent sorting procedure, as they independently sort on market capitalization, the annual change in total assets, and the quarterly return on equity.

It should be noted that there is not always clear guidance on which additional sort one should pick to include in 2x3x3 sorts. In the case of the Q factor model, there are theoretical arguments why the ROE and IA factors should be orthogonalized: the negative relation between investment and cost of capital is conditional on return on equity. In addition, the positive relation between return on equity and cost of capital is conditional on the level of investment. Hence, Hou et al. (2015) have a rationale to use the 2x3x3 sorting methodology. However, there is no theoretical guidance on how to construct FF-factor or DHS-factors using a 2x3x3 sort, or guidance which additional characteristic should be added in the 2x3x3 sort. This additional dimension leads to another degree of freedom, where the researcher has a wide range of options to select from.

Also note that using the 2x3x3 sorting methodology may lead to sparse or even empty portfolios. We construct an additional 2048 versions of each factor, using 2x3x3 sorting instead of 2x3 sorting, with the HML factor as the second sorting characteristic when constructing the FF and DHS factors. When an underlying portfolio of one of the leg is empty (say big-high-high), we consider the whole factor leg as missing. In table A.1, we count for how many sets of construction choices (out of 2048) we obtain empty portfolios, in at least one month. When we construct 2x3x3 factors using 30-70 breakpoints, we find that most factors have no empty portfolio. The only exception is RMW, CMA, and PEAD, with 36, 12, and 32 sets of choices (out of 2048) missing data. Using 80-20 breakpoints limits the cross-section, and allows the occurrence of empty portfolios to increase. For RMW, we find that emptiness occurs in 400 sets of choices. As we mentioned before, a shortcoming of the independent sorting is that it may cause sparse portfolios. Panels C and D consider independent and dependent sorting, respectively. With 2x3x3 sorting, we indeed find that sparsity comes from the independent sorts. For dependent sorts, we never find empty portfolios. In Panel E and F, we show the sparsity for NYSE and NAN portfolios. Using NAN breakpoints creates a wider universe to select stocks from, making empty portfolios less likely when compared to using NYSE breakpoints. For NYSE we indeed find empty portfolios, ranging from 136 to 292 construction sets, whereas we do not find empty portfolios when NAN breakpoints are used.

Next, we consider the impact of 2x3x3 sorting jointly, with all other construction choices, on the Sharpe ratios of the Q factors (separately) and all factors (together). These results are shown in figure A.1 and A.2, respectively. We find that using the 2x3x3 sorts increases the Sharpe ratio across all construction choices. For example, using NAN breakpoints and 2x3 sorts yields an average Sharpe ratio of 0.82, whereas 2x3x3 sorting yields a Sharpe ratio of 0.98 for Q-factors. Hence, the 2x3x3 sorting methodology is a construction choice that is able to consistently increase the risk-adjusted return of factors.

Finally, table A.2 shows the amount of firms that are included in the Q-factor 2x3x3 portfolios. The 2x3x3 sorting methodology may stratify stocks into smaller segments where more extreme positions are overweighted. For example, the Big-High-High portfolio receives a weight of 1/6 in the High portfolios of the second and third characteristic. We find that the extreme portfolios typically contain less than 100 stocks. For example, Big-Low-Low contains an average of 51 stocks when excluding microcaps, and 34 when we use NYSE breakpoints.

## 3.12 Net Returns

Gross returns do not represent what is actually achievable by investors. To evaluate net returns, we estimate individual stock-level transaction costs as in Detzel et al. (2021) by using the estimation procedure from Hasbrouck (2009). Appendix 3.12.1 explains how we estimate turnover and Appendix 3.12.2 explains the estimation of transaction costs. Appendix 3.12.3 examines net Sharpe ratios. Appendix 3.12.4 focuses on model comparison with net returns.

### 3.12.1 Turnover

We estimate portfolio turnover for each factor. The turnover of an individual stock at time $t$ ($TO_{i,t}$) is calculated by taking the absolute value of the difference between the portfolio weight at the start of the month ($W_{i,t}$) and the weight at the end of the past month ($W_{i,t-1,end}$). The turnover of the long leg of a factor is then defined as:

$$TO_{long,i,t} = \sum_{i=1}^{N_t} |W_{i,t} - W_{i,t-1,end}| \tag{3.8}$$

The turnover of the short-leg is defined in a similar way. The turnover of the long-short factor is defined as the sum of both the long and the short portfolios.

### 3.12.2 Transaction Costs

We estimate transaction costs at the individual stock-level using the procedure of Hasbrouck (2009). This procedure yields effective spreads that highly correlate ($\geq$95%) with those from the high-frequency Trade and Quote (TAQ) database and allows for an estimation of effective spreads for public companies in the CRSP database using their daily price series. The procedure entails estimating transaction costs using a Bayesian-Gibbs sampler on the generalized stock price models of Roll (1984):

$$V_t = V_{t-1} + \epsilon_t \tag{3.9}$$

$$P_t = V_t + cQ_t \tag{3.10}$$

where $V_t$ denotes the log midpoint of the prior bid-ask price (the "efficient price"), $P_t$ denotes the log trade price (the "real price"), and $Q_t$ indicates the sign of the last trade of the day. $Q_t$ equals +1 for a buy, and -1 for a sale. $\epsilon_t$ is a random public shock to the efficient price $V_t$, and $c$ is the effective one-way transaction cost. The above equations imply that:

$$\Delta P_t = \Delta c Q_t + \epsilon_t \tag{3.11}$$

Hasbrouck (2009) estimates $c$ using an augmented version of the equation:

$$\Delta P_t = \Delta c Q_t + \beta R_{m,t} + \epsilon_t \quad (3.12)$$

where $R_{m,t}$ denotes the market return. Because the procedure from Hasbrouck (2009) yields missing observations, we impute observations by following the matching procedure from Detzel et al. (2021). First, for each stocks $i$ on month $t$ we compute:

$$M_i = \sqrt{(rank(ME_i) - rank(ME_j))^2 + (rank(IVOL_i) - rank(IVOL_j))^2} \quad (3.13)$$

where $ME_i$ is the market capitalization and $IVOL_i$ is the 1-year idiosyncratic volatility estimate for firm $i$. If the transaction cost is missing for stock $i$ on month $t$, we impute the transaction cost by finding the stock $j$ that has the smallest difference between $M_i$ and $M_j$, and using the transaction cost estimate of stock $j$.

For the long-leg, we compute portfolio-level effective spreads as follows:

$$TC_{long,t} = \sum_{i=1}^{N_t} |W_{i,t} - W_{i,t-1,end}| * c_{i,t} \quad (3.14)$$

where $c_{i,t}$ denotes the estimated transaction cost for stock $i$ in period $t$. Transaction costs for the short-leg is defined similarly. Portfolio transaction costs for the long-short portfolios are equal to the sum of the transaction costs of each leg.

### 3.12.3 Net Sharpe Ratios

Figure A.3 shows the net Sharpe ratio distribution across sets of construction choices for each factor, based on value-weighted (Panel A) and equal-weighted (Panel B) net factor returns. In line with our findings on a gross basis, we also observe large variation in Sharpe ratios on a net basis. Some construction methods, for a given factor, yield negative net Sharpe ratios. The PEAD factor, for example, yields Sharpe ratios between -0.43 and 0.34 using value-weighting. The financing factor yields the highest average net Sharpe ratios, ranging between 0.28 and 0.74 when value-weighting.

Figure A.4 shows annualized maximum net Sharpe ratios by construction choice, averaged over factor models. Using 20-80 breakpoints yields a lower maximum Sharpe ratio (0.14) than 30-70 breakpoints (0.16), which could be explained by 20-80 breakpoints tilting towards extreme (small) stocks, which have higher transaction costs. Likewise, including microcaps yields lower net Sharpe ratios (0.12) than excluding microcaps (0.18). Including a price filter also improves the net Sharpe ratio (0.18 vs 0.12), since this excludes small illiquid stocks with high transaction costs. Furthermore, using value-weighting instead of equal-weighting puts less weight towards microcaps and yields an average Sharpe ratio of 0.20 versus 0.10. With gross returns, we documented that Sharpe ratios are higher when we include microcaps and use equal-weighting. With net returns, we thus find the opposite effect, due to the differential costs involved. Overall, our findings imply that construction choices also materially affect factor performance on a net basis.

### 3.12.4 Model Comparison

Detzel et al. (2021) show that when (transaction) costs are ignored, model comparison based on squared Sharpe ratios favor models with high gross performance, even when trading costs are high. Hence, we also consider net factor returns when reporting maximum Sharpe ratios on an annualized basis. For the mean-variance analysis with transaction costs, we follow the approach in Novy-Marx and Velikov (2016). More specifically, we estimate mean-variance optimal weights by using a long and short version of all the assets in the portfolio, net of transaction costs, subject to a no-shorting constraint on portfolio weights.

Figure A.6 shows the model selection results when we use net factor returns. Maximum Sharpe ratios decline using net returns, compared to the earlier presented gross returns. The value-weighted net FF6 model, with cash profitability, yields a net Sharpe ratio of 0.88. The BS6 model earns a net Sharpe ratio of 0.80. Using net returns, the DHS3 model yields the highest maximum Sharpe ratio (0.93), on average. We find that the average net maximum Sharpe ratio for the BS6 model is smaller than that of the DHS3 model when we use net returns instead of gross returns. We find that the BS6 model has the highest maximum Sharpe ratio in 1.7% of all construction sets, whereas this equals 85.7% for the DHS3 model.

Table A.3 reports the portfolio weights that correspond to the ex-post mean-variance efficient portfolios constructed from the candidate factor models using net returns, where we average the weights across all construction methodologies. Between brackets, we report the standard deviation of the weights, based on our set of 2048 construction methods. The weights are derived by adding a no-shorting constraint in the mean-variance analysis, following Novy-Marx and Velikov (2016). Across all models, we find that the average market weight increases relative to the results based on gross returns. Since transaction costs are incurred, factors are less profitable and more weight is allocated towards the market. Most factor weights decrease due to transaction costs. For example, CMA in the FF5 model decreases from 47.6% (gross) to 28.3% (net). In addition, due to the no-shorting constraint, low weights are allocated to factors with high transaction costs and negative net alphas. One example of such a case is the PEAD factor. It has a net weight of 6.8%, compared to a 58.1% gross weight, and a 9.6% standard deviation. For multiple construction choices, PEAD has a negative net alpha, thereby binding the no-short constraint and consequently receiving zero weight. The net DHS model predominantly consists of the financing factor (55.1%) and the market factor (38.1%).

Regarding the efficient frontier, table A.4 shows the results when we focus on factor returns net of transaction costs. Due to these transaction costs, estimated alphas are closer to zero. Adding BS6 factors to FF models expands the efficient frontier between 0.03% and 0.08% per month, with standard deviations between 0.02% and 0.04%. Hence, there are construction methods for which the added value of the BS6 factors to the FF models is zero. The DHS factors improve FF models between 0.15% and 0.21% per month with standard deviations between 0.06% and

0.07%. Therefore, there are fewer construction methods that reach alphas closer to zero when adding the DHS model compared to the BS6 model. Again, our results imply that construction methods can influence model selection exercises, as shown by the relatively large standard deviations.

Table A.5 presents the results on economic significance using net returns. This exercise provides a more realistic view of the extent to which the investment opportunity set improves. We find that adding the Q4 factors improves the Fama-French models between 4.2% ($FF6_c$) and 15.4% with standard deviations between 4.7% and 10.9%, on average. For multiple construction methods, the Q4 factor adds little to no improvement relative to the FF models. Similarly, the BS6 and DHS factors improve FF factor models less compared to the analysis using gross returns, which is due to these models containing factors with relatively high turnover and transaction costs. Still, adding the BS6 factors to FF5 improves the Sharpe ratio by 42.8%, on average, with a standard deviation of 15.1%. Overall, we find that net economic significance varies due to differences in construction choices.

As a final analysis, we consider net returns for in-sample (Panel A) and out-of-sample (Panel B) estimation in table A.6. It can be seen from Panel A that the 6-factor model with cash profitability is the best model in 20.6% of the cases, compared to 4.6% when using gross settings. Taking transaction costs into account, the BS6 model is no longer the model with the highest win-probability (15.4%). The DHS model has a win-probability of 57.1%. For out-of-sample estimates using net returns, the DHS model is also the model with the largest win-probability (71.2%).

## 3.13 Additional Figures and Tables

Figure A.1: **Construction choices and gross Sharpe ratios for the Q-Factor Model.** This figure shows the impact of construction choices on the Sharpe ratio averaged over factors. Sharpe ratios are computed on a gross-basis and are annualized. The construction choice definitions are the same as in Figure 3. "233" ("23") denotes that the factors are constructed using a 2x3x3 (2x3) sorting methodology. Monthly Q-factor returns are constructed using data from January 1972 to December 2021.

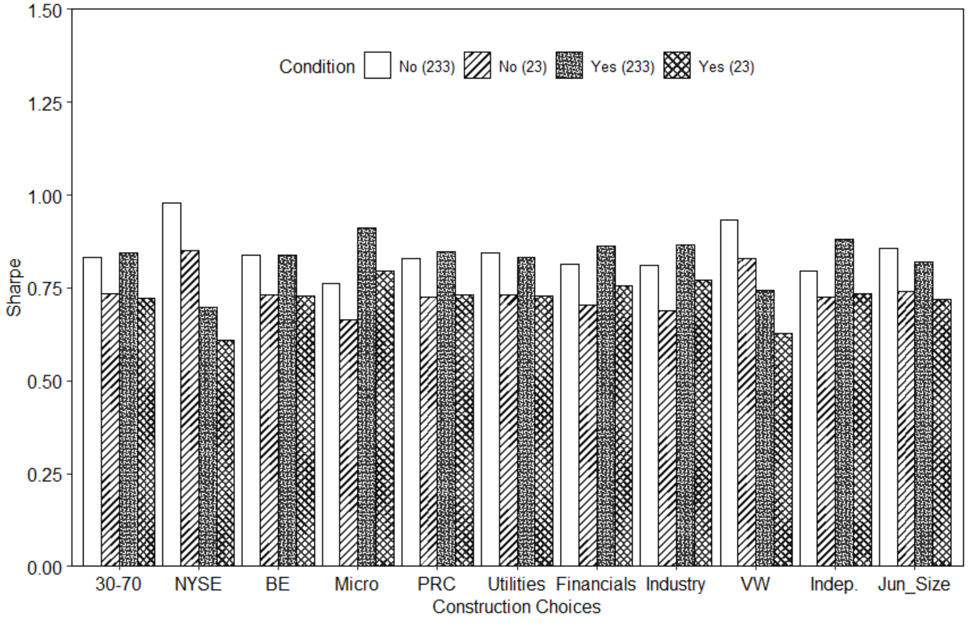


Figure A.2: **Construction choices and gross Sharpe ratios averaged over all factors.** This figure shows the impact of construction choices on the Sharpe ratio averaged over factors. Sharpe ratios are computed on a gross-basis and are annualized. The construction choice definitions are the same as in Figure 3. "233" ("23") denotes that the factors are constructed using a 2x3x3 (2x3) sorting methodology. Monthly Q-factor returns are constructed using data from January 1972 to December 2021.

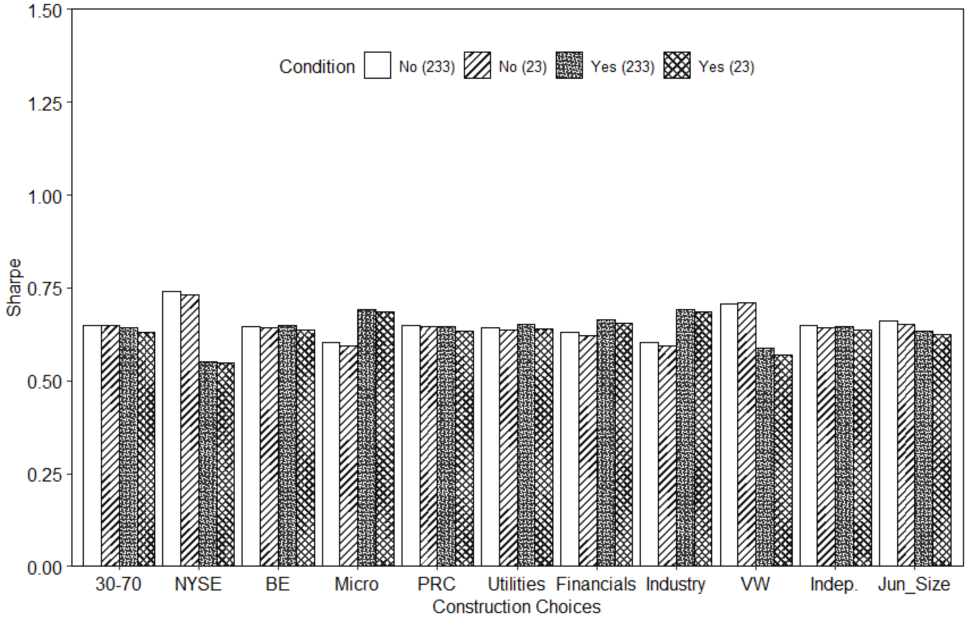
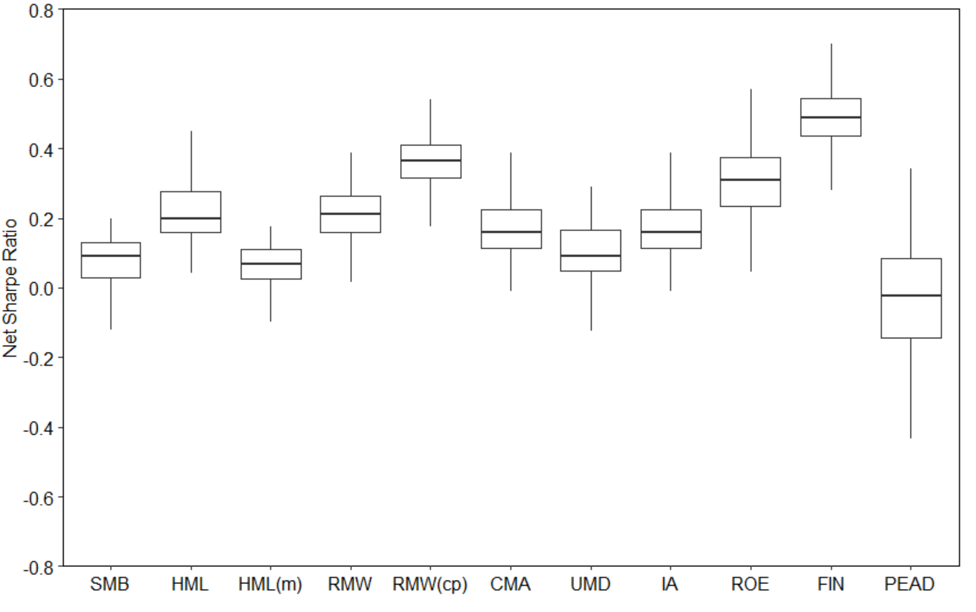
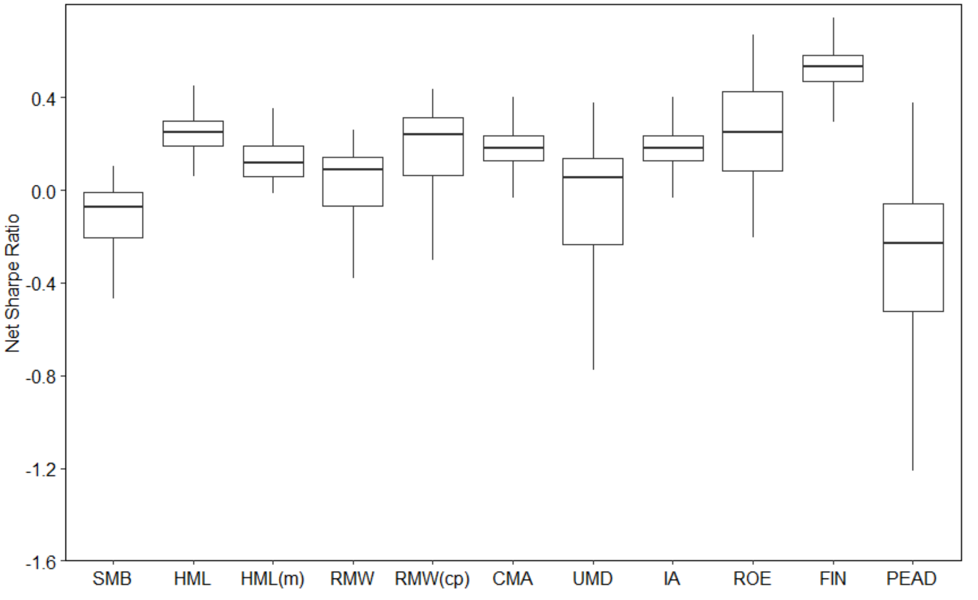


Figure A.3: **Sharpe ratio variation within factors net of transaction costs:** This figure plots the distribution of annualized value-weighted (subfigure A) and equal-weighted (subfigure B) Sharpe ratios for long-short factor returns net of transaction costs, where a factor is constructed $N$ times by using the $N$ different factor construction methods. The black solid line within the box plot shows the median Sharpe ratio. The upper (lower) bound shows the 75th (25th) percentile. The factors and their definitions are from Table 1. The sample, to calculate these Sharpe ratios, runs from January 1972 until December 2021.




(A) Value-Weighted

Figure A.3: **Sharpe ratio variation within factors net of transaction costs – continued.**




(B) Equal-Weighted

Table A.1: **Portfolio sparsity.** This table shows how many construction methods, for a given filter, contains at least one missing month of portfolio returns using a 2x3x3 sorting method. The first sorting characteristic is market capitalization (size), and the second sorting characteristic is the book-to-market ratio (value). The third sorting characteristic is listed in the first column.

| | **Panel A: 30-70** | | | | **Panel B: 80-20** | | | | **Panel C: Independent** | | | | **Panel D: Dependent** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **2nd sort** | | **3rd sort** | | **2nd sort** | | **3rd sort** | | **2nd sort** | | **3rd sort** | | **2nd sort** | | **3rd sort** | |
| Factors | High | Low | High | Low | High | Low | High | Low | High | Low | High | Low | High | Low | High | Low |
| RMW | 36 | 0 | 36 | 0 | 400 | 4 | 392 | 20 | 436 | 4 | 428 | 20 | 0 | 0 | 0 | 0 |
| $RMW_{cp}$ | 0 | 0 | 0 | 0 | 256 | 0 | 240 | 16 | 256 | 0 | 240 | 16 | 0 | 0 | 0 | 0 |
| CMA | 12 | 0 | 12 | 0 | 124 | 0 | 124 | 0 | 136 | 0 | 136 | 0 | 0 | 0 | 0 | 0 |
| MOM | 0 | 0 | 0 | 0 | 136 | 16 | 28 | 124 | 136 | 16 | 28 | 124 | 0 | 0 | 0 | 0 |
| PEAD | 32 | 0 | 32 | 0 | 384 | 48 | 384 | 108 | 416 | 48 | 416 | 108 | 0 | 0 | 0 | 0 |
| FIN | 0 | 0 | 0 | 0 | 252 | 0 | 236 | 16 | 252 | 0 | 236 | 16 | 0 | 0 | 0 | 0 |
| | **Panel E: NYSE** | | | | **Panel F: NAN** | | | | **Panel G: Incl. Micro** | | | | **Panel H: Ex. Micro** | | | |
| | **2nd sort** | | **3rd sort** | | **2nd sort** | | **3rd sort** | | **2nd sort** | | **3rd sort** | | **2nd sort** | | **3rd sort** | |
| Factors | High | Low | High | Low | High | Low | High | Low | High | Low | High | Low | High | Low | High | Low |
| RMW | 292 | 4 | 292 | 4 | 0 | 0 | 0 | 0 | 220 | 0 | 212 | 16 | 216 | 4 | 216 | 4 |
| $RMW_{cp}$ | 208 | 0 | 208 | 0 | 0 | 0 | 0 | 0 | 148 | 0 | 132 | 16 | 108 | 0 | 108 | 0 |
| CMA | 136 | 0 | 136 | 0 | 0 | 0 | 0 | 0 | 32 | 0 | 32 | 0 | 104 | 0 | 104 | 0 |
| MOM | 128 | 16 | 28 | 116 | 0 | 0 | 0 | 0 | 72 | 16 | 8 | 68 | 64 | 0 | 20 | 56 |
| PEAD | 288 | 48 | 288 | 48 | 0 | 0 | 0 | 0 | 224 | 0 | 224 | 40 | 192 | 48 | 192 | 68 |
| FIN | 236 | 0 | 236 | 0 | 0 | 0 | 0 | 0 | 128 | 0 | 112 | 16 | 124 | 0 | 124 | 0 |

Figure A.4: **Construction choices and net Sharpe ratios.** This figure shows the impact of construction choices on the Sharpe ratio averaged over factors. Net Sharpe ratios are annualized. The construction choice definitions are the same as in Figure 3. Monthly factor returns are constructed using data from January 1972 to December 2021.

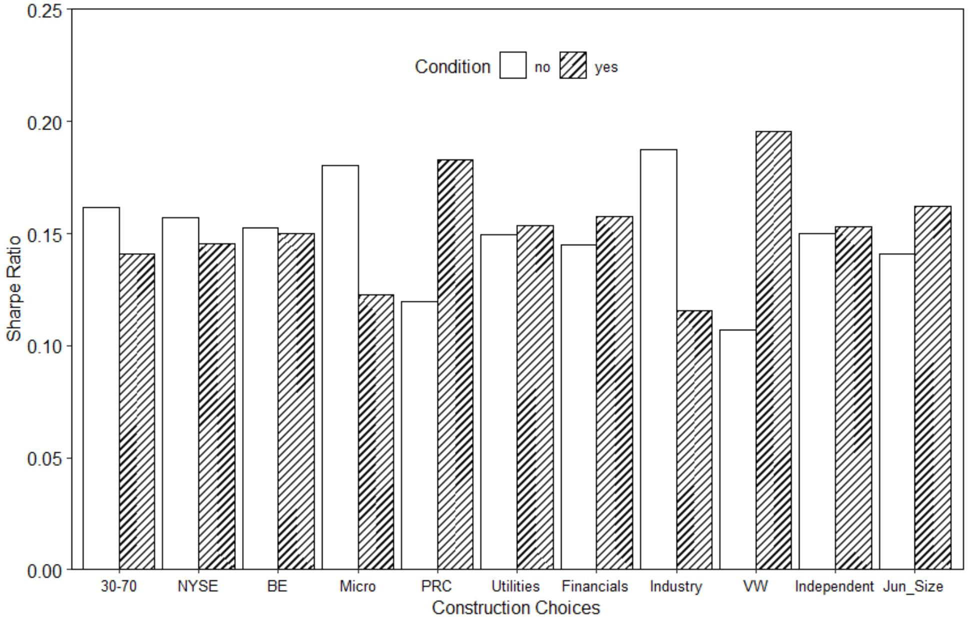


Figure A.5: **Reducing non-standard errors.** This figure shows the non-standard errors when we impose microcaps exclusion ('ExMicro'), NYSE breakpoints ('NYSE') or value-weighting ('VW'). It also reports non-standard errors for combined restrictions. The non-standard errors, shown on the y-axis, are averaged across the eleven factors.

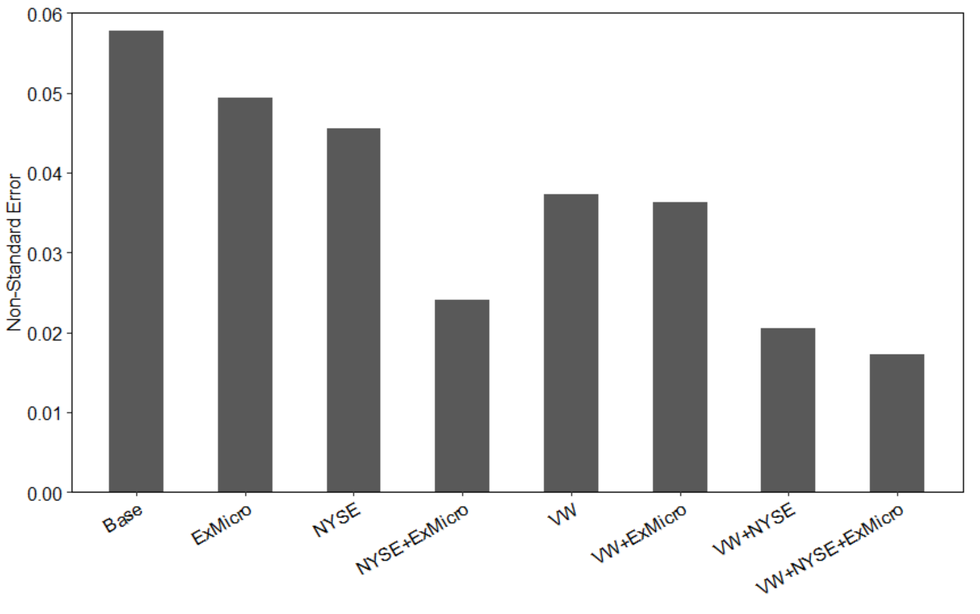


Figure A.6: **Selecting factor models using net returns.** This figure shows the maximum gross Sharpe ratio (annualized) from the factors from the factor models listed on the horizontal axis. The white bar shows the maximum Sharpe ratio obtained by using equal weighted factor returns. The dashed bar shows the maximum Sharpe ratio using value weighted factor returns. The error plot shows the variation in the maximum Sharpe ratios for a given factor model, across construction choices. The data runs from January 1972 until December 2021. The figure shows the results using net returns, taking transaction costs into account.

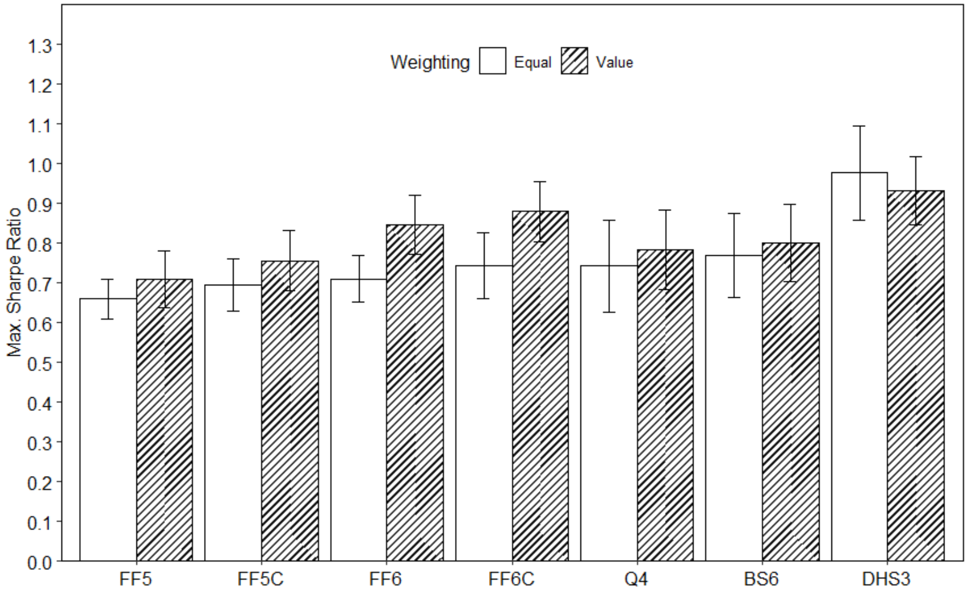


Table A.2: **Firms in Q factor portfolios.** This table shows the average number of positions for Q-factor portfolios, averaged across all 2048 construction methods. B and S denotes the Big and Small portfolio, respectively. H and L (second letter) denotes High and Low for the IA characteristic, whereas the third letter denotes High or Low for the ROE characteristic. The ROE and IA factor are constructed using data from January 1972 to December 2021.

| Choice | BHH | BHN | BHL | SHH | SHN | SHL | BLH | BLN | BLL | SLH | SLN | SLL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ex. Micro | 72 | 103 | 50 | 68 | 118 | 78 | 52 | 92 | 51 | 51 | 105 | 88 |
| Incl. Micro | 90 | 126 | 59 | 104 | 185 | 142 | 58 | 103 | 57 | 82 | 167 | 180 |
| Dep. | 65 | 117 | 65 | 89 | 159 | 103 | 60 | 109 | 60 | 77 | 138 | 93 |
| Ind. | 97 | 112 | 44 | 83 | 145 | 117 | 50 | 85 | 48 | 56 | 134 | 176 |
| 20-80 | 51 | 114 | 35 | 56 | 152 | 77 | 34 | 94 | 33 | 43 | 137 | 96 |
| 30-70 | 111 | 115 | 74 | 116 | 151 | 143 | 76 | 100 | 75 | 90 | 135 | 172 |
| NAN | 105 | 149 | 70 | 71 | 128 | 86 | 75 | 135 | 73 | 66 | 140 | 121 |
| NYSE | 57 | 79 | 39 | 101 | 175 | 134 | 35 | 60 | 34 | 67 | 132 | 147 |

Table A.3: **Mean-variance efficient portfolio weights using net returns.** This table shows the optimal weights that a mean-variance efficient investor would allocate to factors within a factor model, averaged over our set of possible construction methodologies. Within brackets, we show the standard deviation of the optimal weights that occur within our set of possible construction methods. The table shows the weights using factor returns net of transaction costs. The sample period is from January 1972 to December 2021.

| | Mkt | SMB | HML | RMW | CMA | UMD | $RMW_{cp}$ | IA | ROE | $HML_d$ | FIN | PEAD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FF5 | 33.4 | 1.4 | 19.2 | 17.6 | 28.3 | | | | | | | |
| | (6.0) | (2.6) | (17.2) | (11.9) | (12.5) | | | | | | | |
| FF6 | 30.3 | 1.5 | 21.6 | 15.9 | 20.8 | 10.0 | | | | | | |
| | (6.1) | (2.5) | (15.1) | (10.5) | (11.7) | (5.9) | | | | | | |
| FF5C | 28.7 | 3.7 | 13.3 | | 35.4 | | 19.0 | | | | | |
| | (6.5) | (4.3) | (13.5) | | (16.1) | | (10.3) | | | | | |
| FF6C | 27.0 | 3.5 | 15.4 | | 31.9 | 14.9 | 7.3 | | | | | |
| | (6.7) | (4.1) | (12.6) | | (14.5) | (9.8) | (4.9) | | | | | |
| Q4 | 28.3 | 3.4 | | | | | | 40.2 | 28.2 | | | |
| | (4.7) | (3.6) | | | | | | (6.6) | (7.7) | | | |
| BS6 | 25.9 | 2.8 | | | | | 5.9 | 29.6 | 26.0 | 9.8 | | |
| | (3.7) | (3.2) | | | | | (5.6) | (9.3) | (6.8) | (6.8) | | |
| DHS3 | 38.1 | | | | | | | | | | 55.1 | 6.8 |
| | (5.1) | | | | | | | | | | (8.0) | (9.6) |

Table A.4: **Frontier expansion using net returns.** This table reports the intercepts obtained from the regression $MVE_{M1UM0,t} = \alpha + \beta MVE_{M0,t} + \epsilon_t$. M0 is the "base" model, which is augmented to model $M1UM0$ by adding the factors of $M1$ to $M0$. $MVE_{M1UM0,t}$ is the corresponding mean-variance efficient portfolio obtained from the union of factors of $M1$ and $M0$. $MVE_{M0,t}$ is the mean-variance efficient portfolio of the factors from model $M0$. The t-statistics, reported within parentheses, are heteroskedasticity robust. Within brackets, we report the cross-sectional standard deviation of alpha. The table reports the results using net returns, taking transaction costs into account. The data runs from January 1972 until December 2021.

| **Net** | **Union Model (M1)** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Base Model (M0)** | FF5 | FF5$_c$ | FF6 | FF6$_c$ | Q4 | BS6 | DHS3 |
| FF5 | 0.00 | 0.05 | 0.03 | 0.07 | 0.07 | 0.08 | 0.21 |
| | (0.00) | (2.38) | (1.43) | (2.82) | (2.71) | (2.98) | (4.73) |
| | [0.00] | [0.03] | [0.03] | [0.04] | [0.04] | [0.04] | [0.06] |
| FF5$_C$ | 0.00 | 0.00 | 0.02 | 0.02 | 0.04 | 0.05 | 0.16 |
| | (0.01) | (0.00) | (1.29) | (1.29) | (1.80) | (2.28) | (4.21) |
| | [0.00] | [0.00] | [0.02] | [0.02] | [0.03] | [0.03] | [0.07] |
| FF6 | 0.00 | 0.04 | 0.00 | 0.04 | 0.05 | 0.05 | 0.19 |
| | (0.00) | (2.35) | (0.00) | (2.35) | (2.19) | (2.35) | (4.62) |
| | [0.00] | [0.02] | [0.00] | [0.02] | [0.04] | [0.03] | [0.06] |
| FF6$_c$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.03 | 0.15 |
| | (0.01) | (0.00) | (0.01) | (0.00) | (1.31) | (1.53) | (4.11) |
| | [0.00] | [0.00] | [0.00] | [0.00] | [0.02] | [0.02] | [0.06] |
| Q4 | 0.02 | 0.04 | 0.03 | 0.05 | 0.00 | 0.02 | 0.18 |
| | (1.18) | (2.03) | (1.49) | (2.33) | (0.00) | (1.07) | (4.38) |
| | [0.02] | [0.03] | [0.02] | [0.03] | [0.00] | [0.01] | [0.05] |
| BS6 | 0.01 | 0.04 | 0.01 | 0.04 | 0.00 | 0.00 | 0.16 |
| | (1.01) | (2.01) | (1.01) | (2.01) | (0.00) | (0.00) | (4.31) |
| | [0.02] | [0.03] | [0.02] | [0.03] | [0.00] | [0.00] | [0.05] |
| DHS | 0.02 | 0.03 | 0.03 | 0.04 | 0.04 | 0.04 | 0.00 |
| | (0.88) | (1.54) | (1.48) | (2.00) | (1.82) | (2.06) | (0.00) |
| | [0.02] | [0.03] | [0.02] | [0.03] | [0.03] | [0.03] | [0.00] |

Table A.5: **Economic significance using net returns.** This table reports the increase in the maximum Sharpe ratio of the augmented model $M1UM0, t$ relative to the base model $M0$, to quantify the economic significance: $\Delta\%Sh(M0, M1) = Sh(M0, M1)/Sh(M0) - 1$. The table reports the results using net returns, taking transaction costs into account. The standard deviation of the increase in Sharpe, across construction methods, is reported. The data runs from January 1972 until December 2021.

| **Net** | **Union Model (M1)** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Base Model (M0)** | FF5 | $FF5_c$ | FF6 | $FF6_c$ | Q4 | BS6 | DHS3 |
| FF5 | | 13.8 | 5.9 | 18.7 | 15.4 | 17.9 | 42.8 |
| | | (9.5) | (4.6) | (11.1) | (10.9) | (10.3) | (15.1) |
| $FF5_c$ | 0.0 | | 4.2 | 4.2 | 6.8 | 9.4 | 29.4 |
| | (0.1) | | (3.9) | (3.9) | (7.0) | (6.6) | (12.7) |
| FF6 | 0.0 | 11.9 | | 11.9 | 10.2 | 11.3 | 37.3 |
| | (0.0) | (8.1) | | (8.1) | (8.3) | (7.8) | (13.0) |
| $FF6_c$ | 0.0 | 0.0 | 0.0 | | 4.2 | 5.0 | 26.2 |
| | (0.1) | (0.0) | (0.1) | | (4.7) | (4.7) | (11.2) |
| Q4 | 3.6 | 9.2 | 4.9 | 11.2 | | 3.1 | 32.1 |
| | (3.7) | (7.5) | (3.7) | (8.2) | | (3.1) | (11.1) |
| BS6 | 2.8 | 8.7 | 2.8 | 8.7 | 0.0 | | 29.5 |
| | (3.3) | (7.7) | (3.3) | (7.7) | (0.0) | | (9.9) |
| DHS3 | 2.1 | 5.1 | 4.0 | 6.9 | 5.0 | 6.2 | |
| | (2.9) | (5.4) | (3.3) | (5.8) | (4.3) | (4.4) | |

Table A.6: **In-sample and out-of-sample Sharpe ratios using net returns.** This table reports the percentage of bootstrap simulations where the maximum Sharpe ratio of the model in the row exceeds that of the model in the column, averaged across construction methodologies. We use the factor models listed in Table 1. "SR" reports the maximum Sharpe ratio of the row model, averaged across construction methodologies. $\sigma(SR)$ reports the standard deviation of the maximum Sharpe ratio of the row model. "Best" reports the estimated probability that the row model produces the highest squared Sharpe ratio among all models in the run, averaged over construction methods. $\sigma(Best)$ reports the corresponding standard deviation. Panel A presents the in-sample estimates and Panel B shows the out-of-sample estimates using net returns. The estimates are based on 100.000 in-sample and out-of-sample simulation runs. Each simulation run splits the 600 sample months, running from January 1972 until December 2021, into 300 adjacent pair-months. The run randomly draws a sample of pairs (with replacement). The in-sample simulation randomly draws one month from each pair within a run. The remaining months form the out-of-sample. The in-sample observations are used to calculate in-sample Sharpe ratios and portfolio weights. The in-sample portfolio weights are applied to the out-of-sample returns to produce an out-of-sample Sharpe ratio estimate.

**Panel A: In-sample estimates (net returns)**

| | FF5 | FF6 | FF5$_c$ | FF6$_c$ | Q4 | BS6 | DHS | Best | $\sigma(Best)$ | SR | $\sigma(SR)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| FF5 | 0.0 | 3.4 | 0.0 | 1.8 | 30.5 | 14.6 | 13.2 | 0.00 | 0.00 | 0.78 | 0.07 |
| FF6 | 83.7 | 0.0 | 56.9 | 0.0 | 56.7 | 38.5 | 26.6 | 0.00 | 0.00 | 0.87 | 0.10 |
| FF5$_c$ | 76.2 | 33.0 | 0.0 | 4.2 | 50.6 | 25.9 | 21.4 | 0.57 | 1.31 | 0.84 | 0.09 |
| FF6$_c$ | 88.2 | 73.5 | 83.0 | 0.0 | 71.7 | 54.1 | 35.4 | 20.61 | 19.88 | 0.93 | 0.12 |
| Q4 | 69.0 | 42.9 | 49.1 | 28.0 | 0.0 | 0.0 | 21.3 | 0.00 | 0.00 | 0.85 | 0.11 |
| BS6 | 85.2 | 61.3 | 73.8 | 45.7 | 92.6 | 0.0 | 31.5 | 15.45 | 10.73 | 0.91 | 0.10 |
| DHS | 86.8 | 73.4 | 78.6 | 64.6 | 78.7 | 68.5 | 0.0 | 57.12 | 21.33 | 1.00 | 0.11 |

**Panel B: Out-of-sample estimates (net returns)**

| | FF5 | FF6 | FF5$_c$ | FF6$_c$ | Q4 | BS6 | DHS | Best | $\sigma(Best)$ | SR | $\sigma(SR)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| FF5 | 0.0 | 10.3 | 23.4 | 13.4 | 27.9 | 26.5 | 8.0 | 0.29 | 0.43 | 0.62 | 0.07 |
| FF6 | 76.8 | 0.0 | 65.5 | 24.6 | 52.8 | 50.4 | 18.7 | 4.66 | 5.32 | 0.72 | 0.10 |
| FF5$_c$ | 52.8 | 24.4 | 0.0 | 10.3 | 34.4 | 29.3 | 10.0 | 0.36 | 0.60 | 0.65 | 0.08 |
| FF6$_c$ | 76.5 | 49.0 | 77.0 | 0.0 | 57.8 | 56.0 | 21.0 | 8.92 | 10.47 | 0.74 | 0.11 |
| Q4 | 71.5 | 46.9 | 65.2 | 41.9 | 0.0 | 41.9 | 15.6 | 5.25 | 5.05 | 0.71 | 0.12 |
| BS6 | 73.4 | 49.5 | 70.4 | 43.8 | 50.8 | 0.0 | 16.1 | 4.86 | 4.40 | 0.71 | 0.11 |
| DHS | 92.0 | 81.3 | 90.0 | 79.0 | 84.4 | 83.9 | 0.0 | 71.21 | 19.04 | 0.92 | 0.11 |

# Chapter 4

# Option gamma and stock returns[1]

## 4.1 Introduction

Since the introduction of exchange-based option trading in 1973, the trading activity of derivatives experienced large growth. Especially in recent years, single stock options have seen exceptional growth. For example, total options volumes were 160% of total share volumes in February 2021, and single stock call volumes are up +400% relative to 2018.[2] A key question is whether option trading is able to affect the price dynamics of underlying assets. Recent anecdotal stock-level evidence indeed suggests so: the rising share prices, and volatility, of GameStop in the beginning of 2021 was partially attributed to retail investors that bought large amount of call options.[3] Option market makers need to purchase shares on the market to hedge themselves to remain delta-neutral. Such hedging behaviour can potentially have a large impact on asset prices.

How aggressively option market makers need to trade stocks in order to remain delta-neutral depends on the gamma of the option. Gamma measures how much the price of an option accelerates when the price of the underlying security changes. When market makers have short gamma exposure, they have to buy stocks when they are rising, and short them when they are falling, thereby amplifying initial price movements and volatility. On the other hand, when market makers have long gamma exposures, the opposite effect occurs: market makers buy stocks when they are falling, and sell when they are rising, thereby acting as a volatility dampener.

Given the growing activity in option markets, a natural question is whether this gamma-related flow is a systematic driver of asset returns. In this paper, we aim to

[1] This chapter is based on Soebhag (2022).
[2] See Goldman Sachs Global Macro Research (february, 2021).
[3] See the Financial Times (2021)

answer this question by studying the cross-sectional implications of the net gamma exposure on future equity returns. Following Barbon and Buraschi (2020), we directly proxy the net gamma exposure ($\Gamma$) of a stock as the gamma-weighted sum of open interest across the options written on that stock. We sort individual stocks into decile portfolios by their net gamma exposure during the previous month and examine the next month return on the resulting portfolios. Stocks in the lowest $\Gamma$ decile generate about 10.44% higher annual returns compared to stocks in the highest $\Gamma$ decile. After controlling for several benchmark models (such as the 5-factor model of Fama-French), we still find that the difference between the risk-adjusted returns on the portfolios with the lowest and highest $\Gamma$ remains negative and highly significant.

Our results are consistent with the hypothesis that risk-averse investors demand additional compensation in the form of higher expected returns to hold stocks with negative net gamma exposure. When the gamma exposure is negative (positive), delta decreases (increases) when the price of the underlying asset increases. Hence market makers that engage in delta-hedging strategies are required to buy (sell) the underlying more aggressively after an increase in the underlying's price. This results into additional positive (negative) market pressure, which increases (decreases) the magnitude of the initial price movement. Thus, the initial price movement is dampened (reinforced) when the net gamma exposure is positive (negative). Hence, the relation between net gamma exposure and volatility is expected to be negative. This relationship also implies that risk-averse investors tend to be averse towards negative net gamma exposure, and demand a compensation to hold such stocks. On the other hand, stocks with positive net gamma exposure are considered as safer assets. In that case, investors are willing to pay higher prices, and accept lower expected returns. We confirm that stocks with a lower net gamma exposure tend to have higher realized volatility in the next month.

To ensure that the differences in returns are driven by the net gamma exposure rather than other stock characteristics, we conduct bivariate portfolio sorts and re-examine the alpha differences. After controlling for almost 20 different well-known stock return predictors, we find that the negative relationship between net gamma exposure and future stock returns remains negative and statistically significant. Furthermore, we also examine the cross-sectional relationship at the individual stock-level using Fama and MacBeth (1973) cross-sectional and panel regressions. Controlling for all predictors jointly, these regressions provide strong evidence for an economically and statistically significant negative relation between the net gamma exposure and future stock returns. We also provide evidence of significant variation in the net gamma exposure premium over time.

We investigate the robustness of our findings. First, we construct a 2-by-3 gamma factor, a la Fama and French (1993), to conduct spanning regressions. We show that the gamma factor is not spanned by well-known factor models. Second, we proxy the net gamma exposure using slightly different alternative definitions, and show that the documented negative relationship remains highly significant. Third, the net gamma exposure also has predictive power on the daily and weekly frequency. Fourth, we find that the net gamma exposure premium is highly significant in the cross-sections of the 1000 largest and the 1000 most liquid stocks in the Center for Research in Security Prices (CRSP) universe. Fifth, our results are robust to changes in data filters. Sixth, we show that the predictive power stems mainly from ATM and OTM options, and options with a maturity beyond one month. Lastly, we show that the negative relationship between the net gamma exposure and next month return remains robust after the controlling of a wide range of option-based predictors.

Our study is related to several streams of the literature. First, there is a growing literature that shows evidence that options play a role in the price discovery process. Hu (2014) shows that option market makers' delta hedge trades to hedge new options positions cause the information reflected in option trading to be impounded into underlying equity prices. Ni, Pearson, and Poteshman (2005) show that on expiration dates the closing prices of stocks with listed options cluster at option strike prices, driven by hedge rebalancing of option market makers. Hendershott and Seasholes (2007) study explicitly the link between non-informational order imbalances (buy minus sell volume) to predict daily stock returns at the market level. Second, a few studies investigate the relationship between gamma imbalances and asset prices. Ni et al. (2021) shows that the net gamma exposure predicts the next day absolute return, and provides a non-informational channel argument. Similar, our study finds that the net gamma exposure negatively predicts the realized volatility in the next month, and that this is driven by hedge rebalancing, rather than option trading on private information. The main difference between our study and Ni et al. (2021) is that our study focus on predicting future equity returns and documenting a risk premium for stocks with negative net gamma exposures. Baltussen, Da, Lammers, and Martens (2021) finds, on the index-level, that the return between the the previous close and 15:30pm positively predicts the return between 15:30pm and market close, driven by hedging demand as measured by the net gamma exposure. Barbon and Buraschi (2020) and Barbon et al. (2021) finds that end-of-the-day predictability interacts with the net gamma exposure. These three studies all focus on intraday returns, whereas we focus on lower frequency returns. Furthermore, these studies do not show the direct effect of the net gamma exposure on returns in the following trading day(s). However, we find that net gamma exposure predicts next-day, next-week, and next-month returns, and hence is not temporary, nor reverting.

The remainder of this paper is structured as follows. We describe the data and variable construction in section 4.2. The empirical results are presented in 4.3. We run a series of robustness tests in section 4.4. In section 4.5 we examine how net gamma exposure affects stock volatility and trading volume. Section 4.6 concludes.

## 4.2 Data and variable definitions

We use data of U.S.-listed options that are written on individual stocks trading on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and National Association of Securities Dealers Automated Quotation (NASDAQ). From OptionMetrics we obtain daily implied volatility, trading volume, open interest and Greeks for each option contract. The option data runs from Jan. 1, 1996 (the first date in the OptionMetrics database) until Dec. 31, 2021. We match the option data to stock return data obtained from CRSP. We only use stocks where the share code equals 10 or 11, and exchange code 1, 2, or 3. Furthermore, we eliminate stocks with a price per share less than 5 dollar and/or stocks with a market capitalization below the 20th NYSE percentile in order to exclude micro-caps from our sample. Accounting variables are obtained from Compustat and matched to our sample.

### 4.2.1 Net gamma exposure:

Let $S_t$ be the value of the underlying asset at time $t$, $K$ the strike price of an option and $C_t$ the price of an option. The delta ($\Delta_t$) of an option $C_t$ is defined as the first derivative of the option price w.r.t the underlying price: $\Delta_t = \frac{\delta C_t}{\delta S_t}$. Option market makers aim to neutralize their exposure to movements in $S_t$ in their option portfolio by engaging in delta-hedging. At time $t$, delta-hedging of an option portfolio requires buying or selling an amount of the underlying equal to $-\Delta_t$. However, $\Delta_t$ is a function of $S_t$. Thus, changes in $S_t$ also changes the value of $\Delta_t$. Hence, delta-hedging requires a dynamic adjustment of the position on the underlying. The extent in which $\Delta_t$ changes when $S_t$ changes is the gamma, $\Gamma_t$, which is the second-order derivative of the option price w.r.t the price of the underlying, i.e. $\Gamma_t = \frac{\delta^2 C_t}{\delta S^2}$. A high absolute value of $\Gamma_t$ implies that $\Delta_t$ is very sensitive to changes to $S_t$, and that the delta-hedger must trade more of the underlying to achieve delta-neutrality.

To estimate the net gamma exposure ($\Gamma$) on a individual-stock level, we follow Baltussen et al. (2021) and Barbon and Buraschi (2020). For a call option (C) on the underlying stock $i$ on day $t$ with strike price $s \in S_t^c$ and maturity $m \in M_t^c$, the $\Gamma_{i,t}$ is computed as:

$$\Gamma_{i,t}^c = \Gamma_{i,s,m,t}^c \times OI_{i,s,m,t}^c \times 100 \times S_t$$

Where $\Gamma_{i,s,m,t}^C$ denotes the option's gamma, $OI_{i,s,m,t}^c$ is the option's open interest, 100 is the adjustment from option contracts to shares and $S_t$ is the price of the underlying. For a put option (P) on the underlying stock $i$ on day $t$ with strike price $s \in S_t^p$ and maturity $m \in M_t^p$, the $\Gamma_{i,t}$ is computed as:

$$\Gamma_{i,t}^p = \Gamma_{i,s,m,t}^p \times OI_{i,s,m,t}^p \times (-100) \times S_t$$

Here we multiply by (-100) as this represents short gamma for option market makers. To compute the aggregated net gamma exposure for stock $i$ at day $t$, we sum over all $\Gamma^c$'s and $\Gamma^p$'s at every strike price and every maturity:

$$\Gamma_{i,t} = \left( \sum_{s \in S^c} \sum_{m \in M^c} \Gamma^c_{i,s,m,t} + \sum_{s \in S^p} \sum_{m \in M^p} \Gamma^p_{i,s,m,t} \right) \times \left( \frac{S_t}{100 \times VOL_{i,t-1}} \right) \tag{4.1}$$

The first term between brackets denotes the amount (in dollars) that option market makers need to trade for a one-dollar change in $S_t$. We facilitate cross-sectional comparison by multiplying this term by the second term: Multiplying by $S_t$ and dividing by 100, and scale by the average dollar trading volume over the last 21 business days. This changes the interpretation to the amount that needs to be hedged for a 1% change in the underlying stock. To limit the impact of outliers, we trim the net gamma exposures at the 1% and 99% each month.

Figure 1 provides an overview of the coverage of our sample relative to the CRSP universe. Data from OptionMetrics is only available from January 1996 on. At the start of 1996, only 45% of number of stocks have valid net gamma exposures available. In terms of total market capitalization, we cover 61% of the CRSP universe in terms of market capitalization in January 1996. Over time, the number of stocks being covered grows, where we obtain over 95% coverage in terms of the number of stocks in 2021, and over 99% in terms of market capitalization.

Figure 2 presents the distribution of the net gamma exposure across stocks for each month in our sample. First, we document significant cross-sectional differences in the net gamma exposure across stocks. On average, the net gamma exposure equals 1.23 for the 75th percentile, whereas it equals 0.05 for the 25th percentile. Second, we also document variation in the net gamma exposure over time. During periods of financial uncertainty and high volatility (such as the Great Financial Crisis or the Dot-com Bubble), the net gamma exposure is lower. Third, we find that most of the stocks have a positive net gamma exposure. In our sample, 21.8% of the stock-month observations have a negative net gamma exposure.

### 4.2.2 Other predictors:

To control for other cross-sectional effects, we construct a wide-range of predictors. The following factor loadings and firm characteristics, that are known to forecast the cross-section of stock returns, are constructed: the size (ME) is defined as the firm size and is measured as the natural logarithm of the market value of equity (which equals the stock price multiplied by the number of shares outstanding in millions of dollars) at the end of month $t$ for each stock $j$.

We compute the following accounting variables: the book-to-market ratio (BM) is computed as the book value of stockholder equity plus deferred taxes and investment tax credit (if available) minus the book value of preferred stock at the end of the last fiscal year, t1, scaled by the market value of equity at the end of December of year t1. Depending on data availability, the redemption, liquidation, or par value (in that order) is used to estimate the book value of preferred stock (Fama & French, 1992). In addition, we compute a monthly version of the B/M ratio, following Asness and Frazzini (2013). Following Hou, Xue, and Zhang (2015), we compute the annual growth rate of total assets, denoted IA, as the change in book assets (Compustat item AT) divided by the lagged AT. The quarterly operating profitability, denoted ROE, is measured by income before extraordinary items (item IBQ) divided by one-quarter-lagged book equity. We compute 1-year net-share issuance (NSI) as the firm's 1-year growth in market equity minus the 1-year equity return (in logs), following Pontiff and Woodgate (2008). The NSI measure excludes cash dividends. The 5-year composite share issuance (CSI) measure is defined as the firm's 5-year growth in market equity, minus the 5-year equity return, in logs, following Daniel and Titman (2006). We compute operating profitability (OP) as revenues minus cost of goods sold, minus selling, general, and administrative expenses minus interest expense all divided by book equity (Fama & French, 2015). We compute cash profitability (CP), following Ball et al. (2016), by defining accruals as the change in accounts receivable from $t-2$ to $t-1$, plus the change in prepaid expenses, minus the change in accounts payable, inventory, deferred revenue, and accrued expenses.

The following trade/price-based variables are constructed: We estimate market beta (MKT) as the market beta of individual stocks using daily returns over the prior year. Likewise, we define total return volatility (VOL) as the volatility of daily returns over the prior year. We define realized volatility (RV) as the volatility of daily returns during month $t$. Momentum (MOM), for each stock in month $t$, is defined as the cumulative return on the stock over the previous 11 months starting two months ago to avoid the short-term reversal effect, that is, momentum is the cumulative return from month $t$–12 to month $t$–2 (Jegadeesh & Titman, 1993). Following Jegadeesh (1990), we define short-term reversal (SREV) for each stock in month $t$ as the return on the stock over the previous month. Following Amihud (2002), for each stock in month $t$, we define illiquidity to be the ratio of the absolute monthly stock return to its dollar trading volume, $ILLIQ_{i,t} = |R_{i,t}|/VOLD_{i,t}$, where $R_{i,t}$ is the return on stock $i$ in month $t$, and $VOLD_{i,t}$ is the monthly trading volume of stock $i$ in dollars. Idiosyncratic volatility ($IVOL$) is calculated as the standard deviation of the daily abnormal return, based on CAPM model, over the past 90 trading days. Following Bali et al. (2011), we measure demand for lottery-like stocks using $MAX$, which is calculated as the average of the five highest daily returns of the stock during the given month $t$. We require a minimum of 15 daily return observations within the given month to calculate $MAX$.

Lastly, we construct option-based predictors. First of all, we measure implied volatility ($IV$) as the open interest weighted implied volatility for all options traded on that day, following Ge et al. (2016). Furthermore, we compute the total call volume relative to the total option volume in month $t$ (CVOL). Lastly, we compute the total outstanding call option open interest relative to the total option open interest (COI).

## 4.3 Empirical Results

In this section, we conduct a wide range of tests to assess the predictive power of the net gamma exposure over future stock returns. First, we conduct univariate portfolio-level analyses. Second, we analyse the persistency of the net gamma exposure on the portfolio-level. Third, we show the average stock - and portfolio characteristics to provide an overview of the composition of net gamma exposure portfolios. Fourth, we conduct bi-variate portfolio sorting and stock-level regressions to control for other characteristics. Fifth, we control for multiple control variables in a multivariate setting. Sixth, we show that the net gamma exposure also negatively predicts extreme returns in the next month. Lastly, we provide evidence that the net gamma exposure premium is significantly time varying.

### 4.3.1 Univariate portfolio-level analysis

In this section, we conduct univariate portfolio-level analysis, where we construct deciles every month by sorting stocks on their net gamma exposure ($\Gamma$). Subsequently, we compute the one month ahead value-weighted returns for each decile to test whether the zero-cost portfolio generates a significant return. The zero-cost portfolio takes a long position in stocks with the lowest net gamma exposure, and a short portfolio in stocks with the highest net gamma exposure during the previous.

Table 2 presents the time-series averages of one-month-ahead excess (risk-adjusted) returns for each decile. Panel A and B uses breakpoints derived from the full CRSP sample and NYSE universe, respectively, to construct decile portfolios. The first column of each panel reports the average net gamma exposure for each decile. Moving from decile L to decile H, the $\Gamma$ increases significantly from -0.01 to 0.04. The zero-cost portfolio has an average net gamma exposure of 0.05 with a t-statistic of 17.04. The second column of each panel reports the average excess returns. We find that the average excess return decreases monotonically from 1.45% to 0.58% (panel A) when moving from the lowest $\Gamma$ decile to the highest $\Gamma$ decile. The average return difference between decile H and L equals -0.87% per month with a t-statistic of -5.29. This suggests that stocks in the lowest $\Gamma$ decile generate, on average, 10.44% higher annual returns compared to stocks in the highest $\Gamma$ decile.

Subsequently, we report the magnitude and statistical significance of risk-adjusted returns estimated from five different factor models: $\alpha_{3FM}$ is the intercept obtained from regressing the excess portfolio returns on the Fama-French 3-factor model augmented with the momentum factor (i); $\alpha_{5F}$ is the alpha relative to the Fama-French 5-factor model (ii); $\alpha_{5FM}$ is the intercept relative to the Fama-French 5-factor model augmented with the momentum factor (iii); $\alpha_{Q5}$ is the alpha relative to the extended Q-factor model of XHZ (iv); $\alpha_{Q5M}$ is the alpha relative to the extended Q-factor model of XHZ augmented by the momentum factor. As shown in the third column of both panels, the $\alpha_{3FM}$ decreases from 66 basis points to -13 basis points per month when moving from decile L to decile H. The alpha spread equals 79 basis points per month (or 9.48% per annum) with a t-statistic of -4.87. We find similar alpha results from alternative factor models with alpha spreads ranging between 79 and 94 basis points per month. After controlling for well-known factor models, the return difference between low $\Gamma$ and high $\Gamma$ stocks remains negative and statistically significant.

The results are in line with the hypothesis that stocks with negative hedging pressure can exacerbate stock volatility, whereas hedging pressure from positive gamma exposure may act as a volatility dampener. Risk-averse investors would demand extra compensation in the form of higher expected returns to hold stocks with a negative $\Gamma$. Stocks with high positive $\Gamma$, on the other hand, are perceived as relatively safer assets, hence investors are willing to pay higher prices for these stocks and accept lower expected returns.

### 4.3.2 Gamma persistency

The significant and negative alpha spreads documented in table 2 are obtained by sorting stocks by their previous' month net gamma exposure, and not by their contemporaneous gamma. Investors will only pay high prices for stocks with positive gamma hedging pressure in the past with the expectation that such pressure is persistent over time. In this section, we present results regarding the persistence of net gamma exposure.

Table 3 shows the persistence by examining the average 1-month and 12-month-ahead portfolio transition matrix for our sample. We show the average probability that a stock in decile $i$ (defined by the rows) in one month will be in decile $j$ (defined by the columns) in the subsequent month 12 months. If there is no persistency in the net gamma exposure, we would expect that 10% of the stocks in decile $i$ remains in the same decile 12 months later.

However, the results suggest the contrary. 42% of the stocks in the lowest net gamma exposure decile in a certain month continues to be in the same month one month later. Likewise, over half of the highest gamma decile remains in the same decile 1-month later. On a 12-month basis, the persistency becomes weaker. Only 17% of the lowest decile gamma stocks remains in the same decile after one year, whereas 29% of the highest decile gamma stocks remains in the same decile. Theoretically, investors would pay higher (lower) prices for stocks with positive (negative) net gamma exposure in the past given that this exposure will persist in the future. Our results indeed suggest that gamma is a persistent characteristic, especially on a short-term basis.

### 4.3.3 Average portfolio characteristics

We examine the average characteristics of stocks with high vs. low gamma stocks based on Fama and MacBeth (1973) regressions. We report the time-series averages of the slope coefficients from the regressions of the gamma exposure on the stock-level characteristics. For each month $t$, we estimate the following specification and nested versions:

$$\Gamma_{i,t} = \gamma_{0,t} + \gamma_t X_{i,t} + \epsilon_{i,t} \tag{4.2}$$

Where $\Gamma_{i,t}$ is the net gamma exposure of stock $i$ in month $t$ and $X_{i,t}$ is a collection of stock-specific variables observable at time $t$ for stock $i$. The cross-sectional regressions are run at a monthly frequency from January 1996 to December 2021. The results are shown in table 4. Column (1) shows that the average slope coefficient on the lagged net gamma exposure is positive and significant, implying that stocks with high (low) net gamma exposure in month $t-1$ tend to have a high (low) net gamma exposure in month $t$ as well, consistent with table 3.

Column (2) indicates that stocks with higher net gamma exposure tend to be stocks with lower market beta. This could be driven by the fact that stocks with net high gamma are relatively low-volatility stocks, which typically tend to be low-beta stocks as well. Column (3) reports that the average slope coefficient on the 1-month realized volatility significantly negative. Hence, high gamma stocks tend to be less volatile during the month relative to low gamma stocks. Likewise, in column (5) we find that high gamma stocks also exhibit a lower implied volatility than low gamma stocks. Intuitively, this is also what we would expect: positive gamma exposure tends to dampen volatility, whereas negative gamma exposure increases volatility. We find no significant relation between illiquidity and the net gamma exposure. This might be due to the fact that net gamma exposure is standardized by stock dollar volume and hence is implicitly accounts for differences in liquidity. Furthermore, we find that book-to-market and return on equity (columns 6 and 7) are not related to net gamma exposure. Furthermore, we find that stocks with higher profitability tend to be stocks with lower net gamma exposure (column 8 and 9). This is in line with the result that firms with high operating profitability tend to earn higher one-month-ahead alpha (Fama & French, 2018). Lastly, we find that stocks with high momentum stocks tend to be stocks with positive gamma exposures.

The last column in table 4 shows that when we include all variables jointly, the cross-sectional relations tend to be weaker or insignificant. We find that market beta, realized volatility, and implied volatility remains statistically significant after controlling for all other variables. In the appendix, table A.1, we also report average characteristics on the portfolio level. The results are consistent with the stock-level characteristics.

### 4.3.4 Bivariate portfolio-level analysis

The negative relation between net gamma exposure and equity returns in the univariate portfolios in table 2 is possibly due to a firm-specific characteristic that is correlated with net gamma exposure and has a significant impact on expected stock returns. This section examines the relation between the net gamma exposure and future stock returns after controlling for a wide set of return predictors.

To this end, we perform conditional bivariate portfolio sorts on the net gamma exposure controlling for: market beta ($MKT$), the log market capitalization ($ME$), the book-to-market ratio ($BM$), operating profitability ($OP$), cash profitability ($CP$), investment ($IA$), net share issuance ($NSI$), composite share issuance ($CSI$), return on equity ($ROE$), momentum ($MOM$), short-term reversal ($REV$), 1-year return volatility ($VOL$), idiosyncratic volatility ($IVOL$), 1-month realized volatility ($RVOL$), illiquidity ($ILQ$), lottery demand ($MAX$), implied volatility ($IV$), call volume ($CVOL$), and call open interest ($COI$).

We control for a cross-sectional predictor by first forming value-weighted decile portfolios based on the cross-sectional predictor. Then, within each decile, we sort stocks into decile portfolios based on the net gamma exposure, i.e. we use a dependent (conditional) sorting methodology. Subsequently, we average the portfolio returns across the ten deciles of the controlling variable to produce decile portfolios with dispersion in net gamma exposure, but with similar levels of the controlling variable.

The results are shown in table 5, where we report the alpha for each decile relative to the Fama-French 5-factor model augmented with the momentum factor. In the last row, we report the high-low spread portfolio. In total, we control for 20 stock characteristics. We find that alpha differences of the high-low portfolio are between 80 and 107 basis points per month, and remains highly significant (all t-values are smaller than -4). These findings suggest that a wide-range of well-known cross-sectional effects are not able to explain the net gamma exposure premium.

### 4.3.5 Stock-level regressions

Up until this point, we tested whether the net gamma exposure is a determinant of the cross-section of future equity returns at the portfolio level. Such analysis has the advantage of being non-parametric. On the other side, the sorting methodology aggregates and loses information. Furthermore, the sorting methodology does not allow for a setting in which we can control for other variables simultaneously.

Hence, we now examine the relationship between the net gamma exposure and expected returns at the stock level using Fama and MacBeth (1973) and panel regressions in table 6. Panel A presents the time-series averages of the slope coefficients from the Fama-Macbeth regressions of one-month ahead stock returns on the net gamma exposure with and without control variables. The slope coefficients allows to determine which variables have non-zero premia. We weight observations by their previous month's market capitalization. This corresponds to using WLS instead of OLS. In Panel B we equally-weight observations. Panel C and D shows the results from panel regressions, with and without market-cap weighting, respectively.

Column (1) in panel A reports the univariate regression results, and indicates a negative and statistically significant relation between net gamma exposure and the cross-section of future equity returns. The average net gamma exposure coefficient equals -18.65 with a Newey-West t-statistic of -3.80. To give this slope coefficient an economic significance, we can use the average values of the net gamma exposure in the decile portfolios from table 2. The average difference in $\Gamma_{i,t}$ between stocks in decile $H$ and $L$ is equal to 0.0479. Hence, a stock that moves from decile $H$ to decile $L$ decreases its net gamma exposure by 0.0479, which increases its expected return by 18.65*0.0479=0.89 basis points per month.

The second column in panel A of table 6 controls for implied volatility, call volume (in %), call open interest (in %), and a range of price-based variables. The average slope on $\Gamma$ remains economically and statistically significant. The third column of Panel A adds accounting variables as control variables. In this specification, the estimated slope coefficient on $\Gamma$ remains negative and statistically significant. The findings in panel A are robust to changes in estimation techniques. In panel B-D, we find that the estimated coefficient is in all cases negative and statistically significant. The most conservative estimate occurs in panel D column (3), where we equally-weight observations in a panel regression, and is statistically significant. A stock that moves from decile $H$ to decile $L$ increases its expected return by 8.18*0.0479=0.39 basis points per month, which is economically large. Our results suggest that the net gamma exposure premium is not subsumed after jointly controlling for multiple variables.

### 4.3.6 Large stock price movements

Stocks with negative gamma exposure require that delta-hedgers buy (sell) additional stocks after an initial increase (decrease). As such, the stock price will increase (decrease) even further and the initial movement may be amplified. When stocks have a positive gamma exposure, the reverse effect occurs: stock price movements are dampened. Hence, one implication of this mechanism is that future extreme (absolute) returns are more likely to occur when the net gamma exposure is negative.

We examine to what extent extreme returns can be predicted by the net gamma exposure of option market makers. We define $I[r_{t+1} \geq X\%]$ as an indicator variable that takes value one when the next month absolute return is larger than X%. We set X to 25%, 50%, and 75%, respectively. We regress each indicator variable on the net gamma exposure using a panel logit model. We use a panel logit model with fixed effects when regressing the indicator variables on the net gamma exposures (and a set of control variables). The results are shown in table 7.

In panel A, we predict the probability that the next month's absolute return is larger than 25%. In column (1) we show the univariate estimate of the net gamma exposure. The slope on $\Gamma$ is negative and statistically significant, implying that higher net gamma exposures are associated with a lower probability of 25% or higher absolute return in the next month. In column (2) we control for momentum, short-term reversal, call volume, and call open interest. We find that our estimate remains statistically significant and negative. Our findings are robust to the inclusion of various accounting control variables, as shown in column (3).

In panel B and C, we predict the probability that the next month's absolute return exceeds 50% and 75%, respectively. We find that net gamma exposures also negatively predicts the probability of exceeding 50% and 75% returns in the next month. Our findings are in line with the hypothesis that higher net gamma exposure dampens volatility, and hence negatively predicts future extreme returns.

### 4.3.7 Time-varying Gamma premium

In this section, we test if the relation between the net gamma exposure and future stock returns is varying over time or state-dependent by plotting the monthly estimates of the net gamma premium over time. Figure 3 plots the six-month moving average of the monthly estimated slope coefficient of the net gamma exposure on the next month return. The grey-shaded area in the plot indicates the NBER recession dates. The net gamma exposure premium is negative on average, but varies over time. We find that premium tends to decrease during periods of financial crises, such as 2008-2009.

In table 8 we regress the premium on a set of macroeconomic variables. In column (1) we regress the gamma premium on the CFNAI indicator variable. We find that decreases in the CFNAI indicator is associated with decrease in the gamma premium. This indicates that the net gamma premium is more negative during periods of decreasing economic activity. Risk-averse investors would demand a higher premium for stocks with a negative net gamma exposure since such stocks are riskier, especially in an economic downturn. In column (2) we regress the gamma premium against the VIX index, but find no relationship between the VIX and the premium. In column (3), we regress the premium on the sentiment index of Baker and Wurgler (2006). We find that lower sentiment decreases the gamma premium. When sentiment turns bearish, risk aversion increases and a higher premium is required on negative net gamma stocks. In column (4), we regress the premium on the financial uncertainty index (FUNC) of Jurado, Ludvigson, and Ng (2015). We find that higher financial uncertainty predicts a more negative gamma premium. When uncertainty in financial conditions increase, risk-averse investors will require a higher premium on the relatively riskier negative net gamma stocks. In column (5) we regress the gamma premium on the CFNAI, VIX, sentiment, and FUNC measures simultaneously. We find that CFNAI and sentiment positively predicts the premium, whereas FUNC predicts the premium negatively. Our results are consistent with the idea that the gamma premium is lower (i.e. higher for net negative gamma stocks) during bad states of the economy. During bad states, stocks with positive gamma exposures are considered as safer assets, and hence risk-averse investors command a lower premium for these stocks. Whereas the opposite occurs for stocks with a negative gamma exposure.

## 4.4 Robustness

We provide multiple robustness tests in this section to corroborate our earlier results. First, we show that our results are robust to alternative research choices. Second, we conduct spanning regressions using well-known factor models. Third, we assess the predictive power of net gamma exposure on higher frequencies. Fourth, the decompose the gamma exposure in several components. Lastly, we expand our set of control variables further with a wide range of option-based predictors.

### 4.4.1 Alternative research choices

In this section, we show that our results remain qualitatively similar under several alternative methodological choices. First, we restrict our analysis to several sub-samples: the top 1000 largest stock in terms of their market capitalization (A), the top 1000 most liquid stocks in terms of the Amihud (2002) measure (B), and the top 1000 stocks with the highest option trading volume (C). Table A.2 in the appendix shows panel regression results for each sub-sample. We find that the net gamma exposure remains a significant and negative predictor in all sub-samples. Second, in all our analysis so far, we always excluded microcaps and imposed a 5 dollar price filter. In table A.3, in the appendix, we show similar results when we include microcaps and impose no price filter. As such, our results are not affected by small and illiquid stocks. Third, we show that our results are not driven by the specific construction and sorting choices of the net gamma exposure. In table A.4 we impose a one-day implementation lag in the net gamma exposure (panel A), and instead of using the end-of-month net gamma exposure, we take the average net gamma exposure in the sorting month (panel B). In both cases, we document a significant negative relationship between the net gamma exposure and the next month stock return. Lastly, in table A.5, we sort on the end of the month gamma exposure, whereby we scale by market capitalization instead of trading volume, following Baltussen et al. (2021). We find that our results remain robust after scaling by market capitalization. Hence, our findings are robust to slightly different definitions of the net gamma exposure.

### 4.4.2 Spanning regressions:

Having shown the role of the net gamma exposure in predicting the cross-sectional variation in individual stock returns, we subsequently construct a factor that captures the returns associated with the net gamma exposure and examine to what extent well-known factor models explain this gamma factor. We form a gamma factor using the $2 \times 3$ portfolio sorting method of Fama and French (1993). At the end of each month, we sort all stocks into two groups based on the market capitalization, with the breakpoint dividing the two groups being the median market capitalization of stocks traded on the NYSE. Next, we independently sort all stocks into three groups based on the net gamma exposure using the 30th and 70th NYSE percentile values of the net gamma exposure. Taking the intersections of the two classifications results in six portfolios. The gamma factor return is the average return on the two value-weighted low gamma portfolios minus the average of the two high gamma portfolios. In a similar manner, we construct the all Fama-French factors, the momentum factor, and the factors of Hou et al. (2015). We exclude microcaps and stocks with a price below 5$ to mitigate the influence of small, and illiquid stocks.

Panel A of table 9 shows the estimates from spanning regressions using long-minus-short factors. We find that the estimated annualized alphas, relative to several well-known factor models, range between 3.11 and 4.64% on an annual basis. The estimated alphas are statistically significant, with t-statistics between 2.26 and 3.21. As such the gamma factor is not spanned by Fama-French factor models and the Q-factor model (augmented by the momentum factor). In Panel B we conduct the spanning regressions using the long leg of the factor. We find that the low gamma leg is not spanned by the long legs of the other factor returns. Estimated alphas of the long gamma leg ranges between 2.18% and 3.31% on an annual basis, with t-statistics ranging between 2.51 and 3.78. In Panel C, we find that the short gamma leg is spanned by the other short legs, yielding insignificant alphas. These results indicate that the gamma factor is not explained by the well-known factors, driven by its long leg.

### 4.4.3 Daily and weekly frequencies

Next, we assess the predictive power of the net gamma exposure on higher frequencies. Table 10 shows the regression estimates using daily returns and weekly returns. Column (1) in panel A reports the univariate regression results, and indicates a negative and statistically significant relation between the net gamma exposure and the next day excess return. The average net gamma exposure coefficient equals -3.13 with a Newey-West t-statistic of -3.02. This estimate is also economically significant. The daily standard deviation of the net gamma exposure equals 0.0167. Hence, an one-standard deviation increase in the net gamma exposure is associated with a 5.5 basis point decrease ($0.0167 \times -3.13$) in the next day's return. The second column (2) controls for the contemporaneous return. The average slope on the net gamma exposure remains economically and statistically significant at the 1% level. Column 3, 4, and 5 incrementally add an interaction between the net gamma exposure and return, implied volatility, call volume, and call open interest. In all specifications, we find that the relationship between net gamma exposure and next day return is negative and statistically significant at the 1% level. Our finding is robust to the inclusion of other control variables, as shown in column (6). In panel B, we also regress the next week return on the net gamma exposure. We, again, document a negative and statistically significant relation. After the inclusion of multiple control variables, this effect remains robust. Thus, our documented effect is also present in higher frequencies.

### 4.4.4 Option moneyness and time to expiration:

We decompose the net gamma exposure in terms of moneyness and in terms of time to expiration. Option gammas are highest when the option is near-the-money. On the other hand, deep in-the-money or deep out-of-money options tend to have low gammas. We classify an option as "near-the-money" whenever the absolute values of the natural log of the ratio of the stock price to the exercise price less than 0.1, following Bali and Hovakimian (2009). When the value exceeds 0.1, a call (put) option is "in-the-money" ("out-the-money"). Vice verse, when this value is below -0.1, a call (put) option is out-the-money (in-the-money). Hence, the net gamma exposure can be decomposed as:

$$\Gamma_{i,t} = \Gamma_{i,t}^{OTM} + \Gamma_{i,t}^{ATM} + \Gamma_{i,t}^{ITM} \tag{4.3}$$

Furthermore, an option is considered as "fast" when it expires during the next month, else it is classified as "slow":

$$\Gamma_{i,t} = \Gamma^{slow} + \Gamma^{fast} \tag{4.4}$$

We report the results in table 11. Column (1) shows the results when we regress the next month excess return on the net gamma exposure, indicating that net gamma exposure negatively predicts future stock returns. In column (2), we decompose the net gamma exposure into the ATM, OTM, ITM components and regress the next month excess return on these components. We find that net gamma exposures from ATM and OTM contract negatively predicts future returns, whereas the predictive power for ITM contracts is weaker. In column (3), we decompose the net gamma exposure into the fast and slow component and regress the next month returns on these components. We find that the slow gamma negatively predicts the next month stock return, whereas the fast gamma component has no predictive power.

### 4.4.5 Controlling for other option-based predictors

In table 6 we control for only three option-based predictors. In this section, we extend our set of option-based control variables to ensure that the net gamma exposure is distinct for other well-known option-based variables. First, we add the difference between the historical realized volatility and at-the-money implied volatility (Goyal & Saretto, 2009). Second, we construct the implied volatility skew, proposed by Xing, Zhang, and Zhao (2010), as the difference between the average of implied volatilities extracted from out-of-the-money put options and the average of implied volatilities extracted from at-the-money call options. The IV skew reflects the investor's concern about future downward movements in underlying asset prices. A higher IV skew indicates a higher probability of large negative jumps in underlying asset prices. Third, we compute the volatility-of-volatility variable (VoV) of Baltussen et al. (2018), which measures uncertainty about risk by the volatility of implied volatility (vol-of-vol). Fourth, we construct the call-put implied volatility spread (CPIV) of Bali and Hovakimian (2009), which is defined as the difference between the average IV from ATM call options and ATM put options. A high call-put implied volatility spread implies that the call option prices exceed the levels implied by the put option prices and the put-call parity. Fifth, we compute the net dollar open interest as in equation 4.1 with gamma being replaced by one. This allows us to control for variation in the net gamma exposure due to open interest and the price of the underlying. Furthermore, we compute the net delta exposure ($\Delta$), as in equation 4.1 with gamma being replaced by the delta. Lastly, we measure trading volume in derivatives relative to the volume in underlying stocks (O/S), following Roll, Schwartz, and Subrahmanyam (2010).

We show the cross-sectional regression results in table 12. In column (1), we present the univariate regression estimate of the net gamma exposure coefficient. This estimate is negative and statistically significant, as we have seen before. In the remaining columns, we subsequently add an option-based predictor as a control variable. In all specifications, we find a negative and statistically significant relationship between the net gamma exposure and future returns. In particular, in column (13), we control for all predictors simultaneously. We find that the negative relationship between the net gamma exposure and the next month return remains statistically significant at the 1%. Furthermore, we find that $IV_{skew}$, CPIV and O/S positively and statistically significantly predict the next month return, whereas volatility-of-volatility negatively predict future returns. Thus, our results indicate that the net gamma exposure negatively predicts future equity returns even after the inclusion of multiple other option-based predictors.

## 4.5 Why is the relationship negative?

Our results suggest that stocks with a negative (positive) net gamma exposure earn a positive (negative) alpha, on average. Why is that? When the gamma exposure is negative (positive), delta decreases (increases) when the price of the underlying asset increases. Hence market makers that engage in delta-hedging strategies are required to buy (sell) the underlying more aggressively after an increase in the underlying's price. This results into additional positive (negative) market pressure, which increases (decreases) the magnitude of the initial price movement. Thus the initial price movement is dampened (reinforced) when the net gamma exposure is positive (negative). Hence, the relation between net gamma exposure and volatility is expected to be negative. This relationship also implies that risk-averse investors also tend to be averse towards negative net gamma exposure, and hence demand additional compensation in the form of higher expected returns to hold such stocks. On the other hand, stocks with positive net gamma exposure are considered safer assets. In that case, investors are willing to pay higher prices, and accept lower expected returns.

To test this relationship, we regress next month's realized volatility on the net gamma exposure. The estimates are shown in table 13. Column (1) in panel A reports the univariate regression results, and indicates a negative and statistically significant relation between net gamma exposure and next month's realized volatility. The average net gamma exposure coefficient equals -12.92 with a Newey-West t-statistic of -3.58. To give this slope coefficient an economic significance, we can use the average values of the net gamma exposure in the decile portfolios from table 2. The average difference in $\Gamma_{i,t}$ between stocks in decile $H$ and $L$ is equal to 0.0479. Hence, a stock that moves from decile $H$ to decile $L$ decreases its net gamma exposure by 0.0479, which increases its monthly realized volatility by 0.62%. The second column in panel A of table 6 controls for implied volatility, call volume (in %), call open interest (in %), and a range of price-based variables. The average slope on $\Gamma$ remains economically and statistically significant. The third column of Panel A adds accounting variables as control variables. In this specification, the estimated slope coefficient on $\Gamma$ remains negative and statistically significant. The findings in panel A are robust to changes in estimation techniques (as shown in panel B-D). Consistent with our hypothesis, the relationship between net gamma exposure and future stock return volatility is negative.

### 4.5.1 Hedging versus private information

We have shown that the net gamma exposure negatively predicts the realized volatility in the next month. We argue that this is due to option market makers that aim to remain delta-neutral, and hence hedge their exposure away, thereby creating additional price pressure. One alternative explanation is option trading based on private information: if investors possess private information and trade on this in the option market, then they would buy (sell) options when they expect stock volatility to increase (decrease). To distinguish between the two different channels, we decompose the net gamma exposure by following Ni et al. (2021): one component of the net gamma exposure is due to positions that already existed $\tau$ days ago and one component that is created between day $t-\tau$ and day $t$:[4]

$$\Gamma(i,t) = \underbrace{\Gamma(i,t-\tau,S_t)}_{\text{"Old positional Gamma"}} + \underbrace{[\Gamma_{i,t} - \Gamma(i,t-\tau,S_t)]}_{\text{"Information Gamma"}} \tag{4.5}$$

$\Gamma_{i,t-j,S_t}$ denotes the net gamma exposure created using the open interest at time $t-\tau$. The second component, called the "information gamma", indicates the change of the net gamma exposure due to changes in open interest. This specification allows us to distinguish hedge re-balancing from private volatility information. Option positions that existed at period $t-\tau$ are not driven by private information that is obtained after period $t-\tau$. Hence, the first component allows to measure the effect of hedge rebalancing on future volatility. This specification is sufficient when we assume that information is short-lived. When this is not the case, we can further decompose the net gamma exposure by noting that the net gamma exposure of the old positions at $t-\tau$ can also be written as:

$$\Gamma(i,t-\tau,S_t) = \underbrace{\Gamma(i,t-\tau,S_{t-\tau})}_{\text{"Old Gamma"}} + \underbrace{[\Gamma(i,t-\tau,S_t) - \Gamma(i,t-\tau,S_{t-\tau})]}_{\text{"Hedging Gamma"}} \tag{4.6}$$

[4] We set $\tau = 5$ following Ni et al. (2021)

The first component indicates the net gamma exposure using positions established at time $t-\tau$, using the stock price at $t-\tau$. The first component equals the change in the net gamma exposure due to changes in the stock price from $S_{t-\tau}$ to $S_t$, which can not come from volatility information acquired by traders between $t-\tau$ and $t$. We use this decomposition to identify whether the effect of the net gamma exposure on volatility is driven by private information or due to hedge re-balancing.

We again regress the next month's realized volatility on the net gamma exposure, and its components. Table 14 shows the estimates. In column (1) of panel A, we show the effect of net gamma exposure on realized volatility, as we have shown before in table 13. In column (2), we regress the realized volatility on the old positional gamma and the information gamma. We find that the coefficient of the old positional gamma is negative and statistically significant (t-stat is -4.57), whereas the coefficient on the information gamma is positive and not statistically significant. In column (3) of panel A, we decompose the net gamma exposure even further. The information gamma coefficient remains statistically insignificant. We find that the old gamma negatively predicts future realized volatility. More important, the coefficient on the hedging gamma is negative and statistically significant (t-value is -2.60). Our results are qualitatively similar in panels B-D, where we use other estimation methods. The findings suggest that the negative relationship between the gamma exposure and realized volatility is not driven by private information, but rather by hedge re-balancing. Thus there is a non-informational channel through which the option markets have a pervasive influence on underlying stock prices.

### 4.5.2 Gamma exposures and earnings announcements

The previous section show that the negative relationship between net gamma exposure and stock returns is not driven by private information. In this section, we provide an additional piece of evidence against trading on private information, by focusing on earnings announcements. Suppose that prior to an earnings announcement, some market participant receives private information that stock returns will be positive and buys at-the-money call options in order to profit from this information. As market makers write these options, their position will have a negative net gamma exposure, thereby amplifying returns around earning announcements.

We test whether the predictive power of net gamma exposure on stock returns is stronger on days around earnings announcements. The results are reported in table 15. The indicator variable $I[Earnings]$ takes value 1 (else 0), in columns 1-3, on days in which there is an earnings announcement (day $t$). In column 4 and 5, $I[Earnings]$ takes value 1 on days $t-1$ until $t+2$. The interaction between the net gamma exposure and $I[Earnings]$ measures the additional effect of net gamma exposure on stock returns around earning announcement days. As we have seen before, net gamma exposure predicts at day $t$ predicts day $t+1$ returns negatively and significantly at the 1% level. However, we find that the interaction term $\Gamma \times I[Earnings]$ is insignificant in all specifications. Hence, the effect of the net gamma exposure on stock returns does not differ significantly around earning announcement days.

### 4.5.3 Future trading volume

Option market makers need to hedge their exposure in order to remain delta-neutral. When gamma becomes larger in absolute value, the option market maker needs to trade more aggressively to achieve delta-neutrality. Hence, one implication of gamma-hedging is that stocks with a high absolute gamma exposure predicts future trading volume positively since the dollar amount that needs to be hedged will increase. We regress the percentage change in stock trading volume on the absolute net gamma exposure. The results are shown in table 16. In panel A of table 16 we show the estimates from Fama-Macbeth (1973) regressions using value-weighted observations. Column (1) shows the univariate regression results. We find that larger (absolute) net gamma exposures positively predicts higher trading volumes in the next month. This estimate is statistically significant, with a t-statistics of 5.69. In column (2) and (3) we include multiple control variables in our estimation. We find that our estimate of the effect of the absolute gamma exposure on trading volume remains robust to the inclusion of control variables. In the remaining panels, we use different estimation methodology. We find that all estimates of the effect of absolute gamma on future trading volume remains positive and statistically significant, consistent with our hypothesis.

## 4.6 Conclusion

In this study, we examine the relation between the net gamma exposure and the cross-section of expected returns over the sample period of January 1996 to December 2021. We document a significant negative relationship between the net gamma exposure in the equity option market and future stock returns. These results are consistent with the hypothesis that stocks with negative hedging pressure can exacerbate stock volatility, whereas positive hedging pressure acts as a volatility dampener. Risk-averse investors would demand extra compensation in the form of higher expected returns to hold stocks with a negative net gamma exposure. Stocks with high positive net gamma exposure, on the other hand, are perceived as relatively safer assets, hence investors are willing to pay higher prices for these stocks and accept lower expected returns.

Our estimates are economically significant. Stocks in the lowest net gamma exposure decile generate, on average, 10.44% higher annual returns compared to stocks in the highest decile. After controlling for well-known factor models, the risk-adjusted return difference remains negative and statistically significant. Furthermore, in bivariate conditional sorts, we find that a wide-range of well-known cross-sectional effects are not able to explain the gamma exposure premium. The results remain robust in a multivariate setting, using stock-level regressions. We also add several other option-based predictors as control variables, and find that the net gamma exposure is distinct from these predictors.

The negative relation between net gamma exposure and future stock also exists in samples with liquid and large stocks. Furthermore, the gamma premium is found to be significantly more negative during economic downturns and periods of high financial uncertainty, compared to non-recessionary periods, indicating the time-varying nature of the gamma premium. Net gamma exposures also negatively predict extreme returns, consistent with the idea that positive gamma hedging acts as a volatility dampener.

Lastly, we examine the mechanism behind the predictability. We show that net gamma exposure negatively predicts future volatility. Hence, stocks with negative gamma exposure tend to be riskier. As such, risk-averse investors require a premium to be compensated for this risk, which explains why we find a negative return-gamma relationship. Furthermore, we find that hedge re-balancing, not trading on private information, is explaining why net gamma exposure is negatively related to future volatility. Hence, the predictability stems from a non-informational channel via which stock options affect stock returns.

## 4.7 References


Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets*, *5*(1), 31–56.

Asness, C., & Frazzini, A. (2013). The devil in hml's details. *The Journal of Portfolio Management*, *39*(4), 49–68.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The journal of Finance*, *61*(4), 1645–1680.

Bali, T. G., Cakici, N., & Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of financial economics*, *99*(2), 427–446.

Bali, T. G., & Hovakimian, A. (2009). Volatility spreads and expected stock returns. *Management Science*, *55*(11), 1797–1812.

Ball, R., Gerakos, J., Linnainmaa, J. T., & Nikolaev, V. (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics*, *121*(1), 28–45.

Baltussen, G., Da, Z., Lammers, S., & Martens, M. (2021). Hedging demand and market intraday momentum. *Journal of Financial Economics*.

Baltussen, G., Van Bekkum, S., & Van Der Grient, B. (2018). Unknown unknowns: uncertainty about risk and stock returns. *Journal of Financial and Quantitative Analysis*, *53*(4), 1615–1651.

Barbon, A., Beckmeyer, H., Buraschi, A., & Moerke, M. (2021). Liquidity provision to leveraged etfs and equity options rebalancing flows: Evidence from end-of-day stock prices. *Swiss Finance Institute Research Paper*(22-40).

Barbon, A., & Buraschi, A. (2020). Gamma fragility.

Daniel, K., & Titman, S. (2006). Market reactions to tangible and intangible information. *The Journal of Finance*, *61*(4), 1605–1643.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, *47*(2), 427–465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, *116*(1), 1–22.

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of financial economics*, *128*(2), 234–252.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, *81*(3), 607–636.

Ge, L., Lin, T.-C., & Pearson, N. D. (2016). Why does the option to stock volume ratio predict stock returns? *Journal of Financial Economics*, *120*(3), 601–622.

Goyal, A., & Saretto, A. (2009). Cross-section of option returns and volatility. *Journal of Financial Economics*, *94*(2), 310–326.

Hendershott, T., & Seasholes, M. S. (2007). Market maker inventories and stock prices. *American Economic Review*, *97*(2), 210–214.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, *28*(3), 650–705.

Hu, J. (2014). Does option trading convey stock price information? *Journal of Financial Economics*, *111*(3), 625–645.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of finance*, *45*(3), 881–898.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, *48*(1), 65–91.

Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *American Economic Review*, *105*(3), 1177–1216.

Ni, S. X., Pearson, N. D., & Poteshman, A. M. (2005). Stock price clustering on option expiration dates. *Journal of Financial Economics*, *78*(1), 49–87.

Ni, S. X., Pearson, N. D., Poteshman, A. M., & White, J. (2021). Does option trading have a pervasive impact on underlying stock prices? *The Review of Financial Studies*, *34*(4), 1952–1986.

Pontiff, J., & Woodgate, A. (2008). Share issuance and cross-sectional returns. *The Journal of Finance*, *63*(2), 921–945.

Roll, R., Schwartz, E., & Subrahmanyam, A. (2010). O/s: The relative trading activity in options and stock. *Journal of Financial Economics*, *96*(1), 1–17.

Soebhag, A. (2022). Option gamma and stock returns. *Available at SSRN 4256259*.

Xing, Y., Zhang, X., & Zhao, R. (2010). What does the individual option volatility smirk tell us about future equity returns? *Journal of Financial and Quantitative Analysis*, *45*(3), 641–662.

## 4.8 Figures & Tables

Table 1: **Descriptive statistics:** This table reports the descriptive statistics of our main variables. The sample consists of stocks listed on NYSE/AMEX/NASDAQ with share code 10 or 11. We exclude stocks with a market capitalization below the 20th NYSE percentile (micro-caps) and prices below $5 as of the portfolio formation. Panel A reports the time-series average of the cross-sectional mean, standard deviation, and quantiles of each variable. Panel B reports the time series average of the cross-sectional correlations of these variables. The sample runs from February 1996 until December 2021.

**Panel A: Cross-sectional summary statistics**

| Variable | Mean | Std | P1 | P25 | Median | P75 | P99 |
|---|---|---|---|---|---|---|---|
| $\Gamma$ | 0.92 | 2.98 | -2.81 | 0.05 | 0.41 | 1.23 | 8.84 |
| IV | 0.47 | 0.19 | 0.19 | 0.33 | 0.43 | 0.56 | 1.08 |
| Call Vol. | 0.64 | 0.20 | 0.08 | 0.52 | 0.65 | 0.78 | 1.00 |
| Call OI | 0.61 | 0.17 | 0.16 | 0.51 | 0.62 | 0.73 | 0.97 |
| Log(Size) | 7.95 | 1.34 | 5.86 | 6.94 | 7.71 | 8.75 | 11.76 |
| RVOL | 2.44 | 1.31 | 0.72 | 1.58 | 2.14 | 2.97 | 6.80 |
| VOL | 8.86 | 10.11 | 1.36 | 3.76 | 6.24 | 11.27 | 37.48 |
| Mom | 18.65 | 51.49 | -55.47 | -8.22 | 10.88 | 33.96 | 194.06 |
| MAX | 3.07 | 1.68 | 0.79 | 1.96 | 2.69 | 3.77 | 8.68 |
| BM | 0.48 | 0.45 | -0.15 | 0.22 | 0.39 | 0.65 | 1.82 |
| ILQ | 0.36 | 0.88 | 0.00 | 0.04 | 0.13 | 0.37 | 3.14 |

**Panel B: Cross-sectional correlations**

| | $\Gamma$ | IV | Call Vol. | Call OI | Log(Size) | RVOL | VOL | MOM | MAX | BM | ILQ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\Gamma$ | - | -0.11 | 0.23 | 0.27 | 0.15 | -0.10 | -0.03 | 0.05 | -0.06 | -0.02 | -0.07 |
| IV | -0.11 | - | 0.00 | -0.03 | -0.19 | 0.62 | 0.64 | 0.00 | 0.57 | -0.07 | 0.25 |
| Call Vol. | 0.23 | 0.00 | - | 0.54 | -0.03 | 0.02 | 0.04 | 0.03 | 0.07 | 0.03 | 0.06 |
| Call OI | 0.27 | -0.03 | 0.54 | - | -0.07 | 0.02 | 0.02 | 0.01 | 0.02 | 0.03 | 0.09 |
| log(Size) | 0.15 | -0.19 | -0.03 | -0.07 | - | -0.14 | -0.14 | 0.02 | -0.13 | -0.06 | -0.16 |
| RVOL | -0.10 | 0.62 | 0.02 | 0.02 | -0.14 | - | 0.60 | 0.00 | 0.87 | -0.07 | 0.17 |
| VOL | -0.03 | 0.64 | 0.04 | 0.02 | -0.14 | 0.60 | - | 0.10 | 0.57 | -0.09 | 0.14 |
| MOM | 0.05 | 0.00 | 0.03 | 0.01 | 0.02 | 0.00 | 0.10 | - | 0.00 | -0.04 | -0.12 |
| MAX | -0.06 | 0.57 | 0.07 | 0.02 | -0.13 | 0.87 | 0.57 | 0.00 | - | -0.07 | 0.16 |
| BM | -0.02 | -0.07 | 0.03 | 0.03 | -0.06 | -0.07 | -0.09 | -0.04 | -0.07 | - | 0.06 |
| ILQ | -0.07 | 0.25 | 0.06 | 0.09 | -0.16 | 0.17 | 0.14 | -0.12 | 0.16 | 0.06 | - |

Figure 1: **Net gamma exposure coverage:** This figure shows the coverage of the OptionMetrics $\Gamma$ data relative to the CRSP sample. The solid black line represents the fraction of stocks with non-missing $\Gamma$ data relative to the number of stocks in the CRSP sample. The solid grey shows the market capitalization of firms with non-missing $\Gamma$ data relative to the market capitalization of the CRSP universe. The sample runs from January 1996 until December 2021.

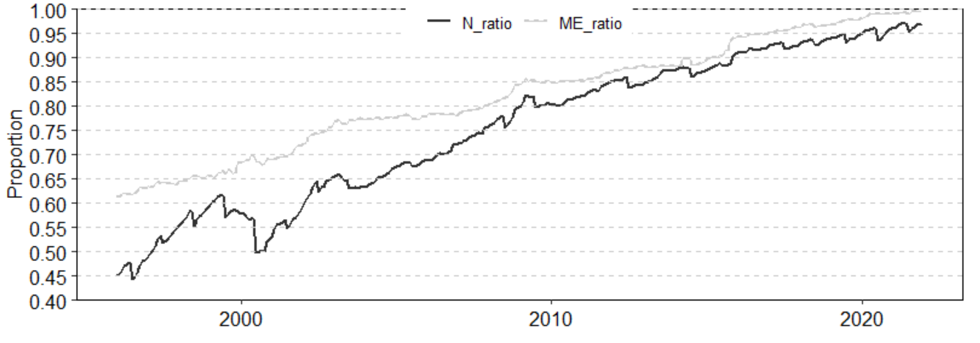


Figure 2: **Net gamma exposure cross-sectional distribution over time:** This figure shows the distribution of the net gamma exposure over time. The 10th, 25th, 50th (median), 75th, and 90th percentiles of the net gamma exposure are shown over time. The sample runs from January 1996 until December 2021.

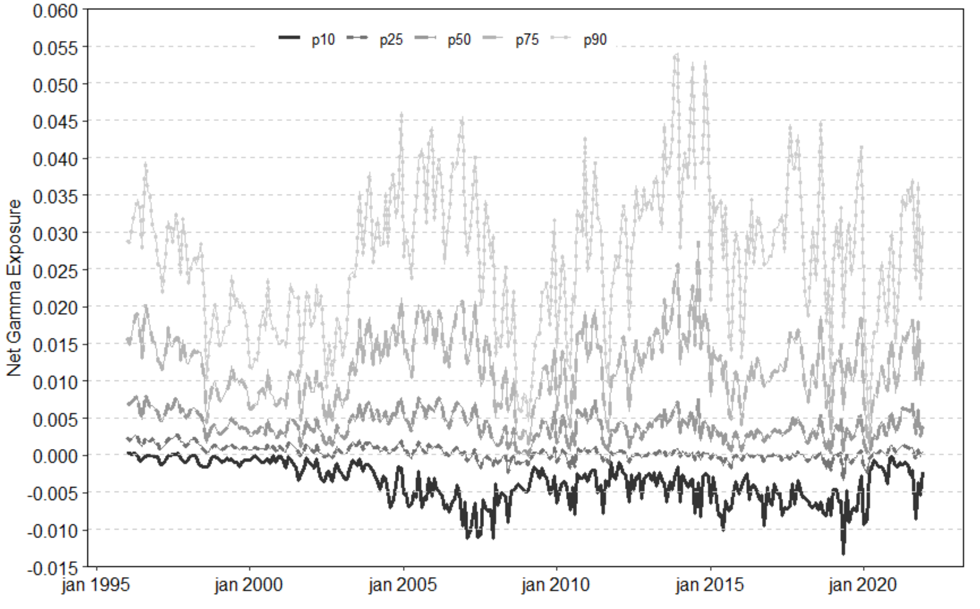


Table 2: **Performance of decile portfolios sorted on the net gamma exposure:** This table reports the performance of decile portfolios formed on the basis of the net gamma exposure ($\Gamma$), which measures the total outstanding gamma divided by the average daily dollar trading volume. At the end of month $t$ we sort stocks into ten portfolios based on their $\Gamma$, and hold this portfolio during month $t+1$. Panel A (B) presents the results for value-weighted portfolios whereby the breakpoints are based on the full sample (NYSE universe). Stocks with prices above $5 and microcaps as of the portfolio formation are excluded. We report the average $\Gamma$, the return ("R") in percentages, the Fama-French-Carhart four-factor alpha ("$\alpha_{3FM}$"), the Fama-French-Carhart five-factor alpha ("$\alpha_{5F}$"), the Fama-French-Carhart six-factor alpha ("$\alpha_{5FM}$"), Hou, Xue, and Zhang's extended q-factor model alpha ("$\alpha_{5Q}$"), and augmented with momentum ("$\alpha_{5QM}$") for each portfolio. The row labeled "L-H" is the self-financing high-minus-low portfolio, which reports the difference in between portfolio H and portfolio L. The sample consists of stocks listed on NYSE/AMEX/NASDAQ for the period between January 1996 and December 2021 with share code 10 or 11. Newey-West t-statistics are reported between parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | **Panel A: Full sample breakpoints** | | | | | | | **Panel B: NYSE-breakpoints** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\Gamma$ | $R$ | $\alpha_{3FM}$ | $\alpha_{5F}$ | $\alpha_{5FM}$ | $\alpha_{Q5}$ | $\alpha_{Q5M}$ | $\Gamma$ | $R$ | $\alpha_{3FM}$ | $\alpha_{5F}$ | $\alpha_{5FM}$ | $\alpha_{Q5}$ | $\alpha_{Q5M}$ |
| L | -0.01*** | 1.45*** | 0.66*** | 0.56*** | 0.66*** | 0.58*** | 0.58*** | -0.01*** | 1.47*** | 0.68*** | 0.58*** | 0.68*** | 0.60*** | 0.60*** |
| | (-11.53) | (5.64) | (4.47) | (4.13) | (4.79) | (3.24) | (3.93) | (-11.46) | (5.76) | (4.61) | (4.27) | (4.88) | (3.29) | (3.94) |
| 2 | -0.00*** | 1.25*** | 0.36*** | 0.27** | 0.35*** | 0.38*** | 0.38*** | -0.00*** | 1.25*** | 0.36*** | 0.26* | 0.35*** | 0.39*** | 0.39*** |
| | (-6.32) | (4.74) | (3.42) | (2.14) | (3.39) | (3.16) | (3.59) | (-5.78) | (4.62) | (3.26) | (1.88) | (3.18) | (3.00) | (3.56) |
| 3 | 0.00*** | 1.09*** | 0.22** | 0.26* | 0.34*** | 0.34** | 0.34** | 0.00*** | 1.10*** | 0.20** | 0.24** | 0.30*** | 0.33*** | 0.33*** |
| | (4.30) | (3.85) | (2.03) | (1.93) | (2.79) | (2.27) | (2.58) | (4.31) | (3.70) | (1.99) | (2.13) | (2.95) | (3.03) | (3.25) |
| 4 | 0.00*** | 1.06*** | 0.11 | 0.13 | 0.17 | 0.27 | 0.27 | 0.00*** | 0.98*** | 0.07 | 0.08 | 0.13 | 0.17 | 0.17 |
| | (11.70) | (3.46) | (0.71) | (0.97) | (1.19) | (2.10) | (2.05) | (10.05) | (3.82) | (0.53) | (0.59) | (0.93) | (1.23) | (1.28) |
| 5 | 0.00*** | 1.03*** | 0.10 | 0.18 | 0.21 | 0.20 | 0.20 | 0.00*** | 1.05*** | 0.10 | 0.20 | 0.21 | 0.15 | 0.15 |
| | (16.30) | (3.50) | (0.86) | (1.17) | (1.47) | (1.41) | (1.47) | (13.60) | (3.14) | (0.70) | (1.23) | (1.31) | (1.09) | (1.10) |
| 6 | 0.01*** | 0.98*** | 0.06 | 0.09 | 0.12 | 0.16 | 0.16 | 0.01*** | 0.99*** | 0.08 | 0.04 | 0.07 | 0.07 | 0.07 |
| | (19.42) | (3.67) | (0.47) | (0.81) | (1.03) | (1.35) | (1.33) | (16.38) | (3.72) | (0.77) | (0.43) | (0.70) | (0.59) | (0.59) |
| 7 | 0.01*** | 0.97*** | 0.06 | 0.09 | 0.09 | 0.08 | 0.08 | 0.01*** | 1.02*** | 0.13 | 0.14 | 0.15 | 0.19* | 0.19* |
| | (20.91) | (3.15) | (0.57) | (0.73) | (0.78) | (0.72) | (0.72) | (18.41) | (3.70) | (1.28) | (1.34) | (1.46) | (1.73) | (1.73) |
| 8 | 0.01*** | 0.94*** | 0.09 | 0.10 | 0.11 | 0.16 | 0.16 | 0.01*** | 0.86*** | -0.00 | 0.01 | 0.00 | -0.03 | -0.03 |
| | (20.91) | (3.58) | (0.94) | (1.05) | (1.07) | (1.63) | (1.64) | (19.45) | (3.21) | (-0.01) | (0.10) | (0.01) | (-0.23) | (-0.23) |
| 9 | 0.02*** | 0.73*** | -0.13 | -0.06 | -0.11 | -0.13 | -0.13 | 0.02*** | 0.77*** | -0.08 | -0.07 | -0.12 | -0.17 | -0.17 |
| | (20.31) | (2.77) | (-1.51) | (-0.64) | (-1.17) | (-1.11) | (-1.21) | (19.47) | (3.09) | (-0.90) | (-0.69) | (-1.31) | (-1.41) | (-1.49) |
| H | 0.04*** | 0.58*** | -0.13 | -0.23*** | -0.27*** | -0.34*** | -0.34*** | 0.04*** | 0.58*** | -0.12 | -0.22*** | -0.25*** | -0.31*** | -0.31*** |
| | (18.88) | (2.73) | (-1.61) | (-2.99) | (-3.65) | (-3.67) | (-3.71) | (18.89) | (2.72) | (-1.23) | (-2.42) | (-2.79) | (-3.01) | (-3.03) |
| H-L | 0.05*** | -0.87*** | -0.79*** | -0.79*** | -0.93*** | -0.92*** | -0.92*** | 0.05*** | -0.88*** | -0.80*** | -0.80*** | -0.94*** | -0.91*** | -0.91*** |
| | (17.04) | (-5.29) | (-4.87) | (-4.42) | (-5.40) | (-3.62) | (-4.33) | (17.54) | (-5.25) | (-4.64) | (-4.31) | (-5.13) | (-3.50) | (-4.13) |

Table 3: **Persistence of the net gamma exposure:** This table presents transition probabilities for net gamma exposure. At each month $t$, all stocks are sorted into deciles based on an ascending ordering of net gamma exposure. The procedure is repeated in month $t+1$ and $t+12$. Portfolio L (H) is the portfolio of stocks with the lowest (highest) net gamma exposure. For each decile in month $t$, the percentage of stocks that also fall into each of the month $t+1$ (panel A) or $t+12$ (panel B) decile is calculated. Table presents the time-series averages of the estimated transition probabilities. Each row corresponds to a different month $t$ portfolio and each column corresponds to a different month $t+1$ or $t+12$ portfolio. The sample runs from February 1996 until December 2021.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Panel A: 1-month transition matrix** | | | | | | | | | | |
| | L | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | H |
| L | 42.17 | 16.02 | 7.13 | 5.77 | 5.55 | 5.45 | 5.37 | 4.71 | 4.38 | 3.85 |
| 2 | 15.74 | 27.57 | 16.90 | 10.87 | 7.87 | 6.03 | 5.01 | 3.98 | 3.06 | 2.18 |
| 3 | 7.29 | 18.53 | 26.95 | 17.44 | 10.75 | 6.64 | 4.67 | 3.39 | 2.26 | 1.40 |
| 4 | 5.98 | 11.20 | 19.29 | 21.58 | 15.48 | 10.48 | 6.91 | 4.47 | 2.85 | 1.53 |
| 5 | 5.94 | 7.80 | 11.50 | 17.84 | 18.43 | 14.97 | 10.27 | 6.84 | 4.20 | 2.24 |
| 6 | 5.55 | 5.97 | 7.25 | 11.05 | 16.84 | 17.77 | 15.11 | 10.60 | 6.61 | 3.27 |
| 7 | 4.95 | 4.75 | 4.64 | 7.16 | 11.80 | 16.79 | 18.57 | 15.70 | 10.66 | 5.16 |
| 8 | 4.86 | 3.83 | 3.14 | 4.52 | 7.36 | 12.10 | 17.54 | 20.44 | 16.99 | 9.43 |
| 9 | 4.28 | 2.76 | 2.07 | 2.56 | 4.16 | 6.97 | 11.60 | 19.91 | 26.20 | 19.83 |
| H | 3.25 | 1.56 | 1.12 | 1.21 | 1.77 | 2.80 | 4.95 | 9.94 | 22.79 | 51.11 |
| **Panel B: 12-month transition matrix** | | | | | | | | | | |
| L | 16.91 | 11.44 | 8.48 | 8.17 | 9.16 | 9.37 | 9.52 | 9.62 | 9.85 | 9.88 |
| 2 | 10.88 | 13.57 | 13.59 | 11.84 | 10.68 | 9.53 | 8.66 | 7.36 | 6.50 | 5.39 |
| 3 | 8.29 | 13.68 | 17.26 | 14.31 | 11.27 | 9.34 | 7.46 | 6.18 | 4.53 | 3.66 |
| 4 | 8.28 | 11.68 | 14.99 | 14.71 | 12.42 | 10.59 | 8.40 | 6.99 | 5.42 | 3.78 |
| 5 | 8.50 | 10.83 | 11.72 | 12.95 | 12.52 | 11.26 | 10.26 | 8.73 | 6.57 | 5.08 |
| 6 | 8.70 | 9.61 | 10.02 | 10.94 | 11.58 | 11.55 | 11.56 | 10.28 | 8.72 | 6.58 |
| 7 | 9.38 | 8.68 | 7.99 | 9.23 | 10.53 | 11.70 | 12.17 | 11.94 | 10.87 | 8.40 |
| 8 | 9.22 | 7.85 | 6.84 | 7.59 | 9.04 | 10.62 | 12.19 | 13.40 | 13.45 | 11.26 |
| 9 | 10.03 | 6.91 | 5.26 | 5.92 | 7.35 | 9.37 | 11.13 | 13.46 | 16.25 | 16.98 |
| H | 9.82 | 5.77 | 3.83 | 4.33 | 5.44 | 6.66 | 8.65 | 12.05 | 17.84 | 28.99 |

Table 4: **Average stock characteristics:** This table reports the estimated slope coefficients from the regressions of the net gamma exposure ($\Gamma$) on stock-level characteristics and risk factors. Panel regressions are run for the following econometric specification and nested versions thereof: $\Gamma_{i,t} = \gamma_{0,t} + \gamma_{1,t} X_{i,t} + \epsilon_{i,t}$. With $\Gamma_{i,t}$ being the net gamma exposure of stock $i$ in month $t$ and $X_{i,t}$ is a collection of stock-specific variables observable at time $t$ stock $i$: The market beta ($MKT$), 1-month realized volatility (RVOL), Amihud's illiquidity (ILQ), Implied volatility (IV), the book-to-market ratio (BM), return on equity (ROE), operating profitability (OP), cash profitability (CP), investments over assets (IA), and 1-year momentum (MOM). Both stock - and time fixed effects are included in the panel regressions. All coefficients are multiplied by 100. The sample spans the period February 1996 to December 2021. Two-way cluster (by stock and date) adjusted t-statistics are given in parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

| Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\Gamma_{t-1}$ | 41.46*** | | | | | | | | | | | 40.69*** |
| | (18.94) | | | | | | | | | | | (18.18) |
| $\beta_{mkt}$ | | -0.69*** | | | | | | | | | | -0.27*** |
| | | (-7.23) | | | | | | | | | | (-3.33) |
| RVOL | | | -31.72*** | | | | | | | | | -27.03*** |
| | | | (-9.11) | | | | | | | | | (-9.03) |
| ILQ | | | | 3.20 | | | | | | | | 2.72 |
| | | | | (1.49) | | | | | | | | (1.17) |
| IV | | | | | -1.00*** | | | | | | | 0.87*** |
| | | | | | (-3.66) | | | | | | | (3.55) |
| BM | | | | | | -0.15 | | | | | | -0.07 |
| | | | | | | (-1.25) | | | | | | (-0.86) |
| ROE | | | | | | | -0.00 | | | | | -0.01 |
| | | | | | | | (-0.24) | | | | | (-0.72) |
| OP | | | | | | | | -0.00*** | | | | -0.00 |
| | | | | | | | | (-5.08) | | | | (-0.57) |
| CP | | | | | | | | | -0.00*** | | | 0.00 |
| | | | | | | | | | (-5.49) | | | (0.51) |
| IA | | | | | | | | | | 0.00 | | 0.02 |
| | | | | | | | | | | (0.23) | | (1.37) |
| MOM | | | | | | | | | | | 0.15** | -0.01 |
| | | | | | | | | | | | (2.58) | (-0.26) |

Table 5: **Bivariate portfolio analysis with conditional sorts:** This table shows the Fama-French, augmented with momentum, 6-factor alpha obtained from conditional bivariate sorts. Stocks are first sorted into deciles based on one control variable, and then stocks within each control variable decile are further sorted into value-weighted deciles based on Γ. The control variables are defined in section 4.2.2. The last row presents the differences in 6-factor alpha between Decile 10 (High) and Decile 1 (Low). Newey-West adjusted t-statistics are given in parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | MKT | ME | BM | $BM_m$ | OP | CP | IA | NSI | CSI | ROE | MOM | SREV | VOL | IVOL | RV | ILQ | MAX | IV | CVOL | COI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| L | 0.71*** | 0.63*** | 0.66*** | 0.70*** | 0.58*** | 0.65*** | 0.72*** | 0.67*** | 0.70*** | 0.64*** | 0.68*** | 0.56*** | 0.69*** | 0.73*** | 0.68*** | 0.67*** | 0.70*** | 0.72*** | 0.71*** | 0.58*** |
| | (4.52) | (3.95) | (4.00) | (3.90) | (3.50) | (4.12) | (4.14) | (4.17) | (3.49) | (4.40) | (3.82) | (3.60) | (4.37) | (4.32) | (3.81) | (3.92) | (3.49) | (4.49) | (4.06) | (3.62) |
| 2 | 0.36*** | 0.37*** | 0.35** | 0.40*** | 0.50*** | 0.48*** | 0.39*** | 0.38*** | 0.29** | 0.32*** | 0.37*** | 0.34*** | 0.52*** | 0.37*** | 0.31** | 0.33*** | 0.29** | 0.35*** | 0.26** | 0.48*** |
| | (2.94) | (2.71) | (2.63) | (2.90) | (3.61) | (3.22) | (2.60) | (3.35) | (2.29) | (3.32) | (3.24) | (2.81) | (3.55) | (2.75) | (2.50) | (2.68) | (2.29) | (2.77) | (2.41) | (3.20) |
| 3 | 0.28** | 0.20 | 0.27** | 0.22** | 0.27** | 0.14 | 0.06 | 0.26* | 0.50*** | 0.31** | 0.27** | 0.45*** | 0.25** | 0.23** | 0.30** | 0.25** | 0.50*** | 0.23* | 0.41*** | 0.22** |
| | (2.26) | (1.48) | (2.42) | (2.46) | (2.43) | (1.40) | (0.52) | (1.91) | (3.38) | (2.42) | (2.24) | (3.18) | (2.19) | (2.38) | (2.65) | (2.02) | (3.38) | (1.76) | (3.13) | (2.35) |
| 4 | 0.28** | 0.34** | 0.32* | 0.17 | 0.17 | 0.38** | 0.13 | 0.18 | 0.05 | 0.17 | 0.27 | 0.08 | 0.16 | 0.03 | 0.19 | 0.25** | 0.05 | 0.15 | 0.35** | 0.36** |
| | (2.10) | (2.47) | (1.81) | (1.09) | (1.12) | (1.97) | (0.93) | (1.18) | (0.30) | (1.21) | (1.43) | (0.98) | (1.58) | (0.18) | (1.42) | (2.18) | (0.30) | (1.23) | (2.46) | (2.72) |
| 5 | 0.09 | 0.05 | 0.22 | 0.15 | 0.19* | 0.16 | 0.22 | 0.13 | 0.29** | 0.25 | 0.10 | 0.01 | 0.30** | 0.41** | -0.01 | 0.21* | 0.29** | 0.25** | 0.30** | 0.43** |
| | (0.74) | (0.53) | (1.56) | (1.11) | (1.78) | (1.30) | (1.37) | (0.80) | (2.11) | (1.57) | (0.97) | (0.07) | (2.24) | (2.56) | (-0.09) | (1.85) | (2.11) | (2.05) | (2.08) | (2.52) |
| 6 | 0.09 | 0.08 | 0.22*** | 0.10 | 0.12 | 0.07 | 0.12 | 0.19** | 0.19 | 0.06 | 0.17* | 0.17 | 0.12 | 0.13 | 0.31*** | -0.01 | 0.19 | 0.02 | 0.06 | -0.01 |
| | (0.93) | (0.70) | (2.76) | (0.89) | (1.21) | (0.87) | (1.58) | (2.04) | (1.56) | (0.71) | (1.67) | (1.48) | (1.09) | (0.84) | (3.90) | (-0.07) | (1.56) | (0.24) | (0.53) | (-0.12) |
| 7 | 0.18** | -0.10 | 0.03 | 0.16* | 0.04 | 0.10 | 0.10 | 0.13 | 0.11 | 0.19 | 0.11 | 0.18** | 0.20** | -0.04 | 0.18** | -0.02 | 0.11 | 0.20** | 0.08 | 0.12 |
| | (2.01) | (-1.11) | (0.24) | (1.69) | (0.31) | (0.82) | (0.80) | (1.37) | (1.18) | (1.26) | (0.91) | (2.26) | (2.05) | (-0.37) | (2.40) | (-0.25) | (1.18) | (2.28) | (0.79) | (0.98) |
| 8 | 0.10 | -0.21** | -0.05 | 0.00 | 0.07 | 0.01 | 0.06 | 0.08 | 0.01 | 0.08 | 0.09 | -0.01 | -0.05 | -0.01 | -0.10 | -0.24** | 0.01 | 0.02 | 0.07 | 0.01 |
| | (1.53) | (-1.79) | (-0.74) | (0.05) | (0.98) | (0.07) | (0.64) | (1.12) | (0.14) | (1.03) | (0.83) | (-0.08) | (-0.44) | (-0.16) | (-0.87) | (-2.14) | (0.14) | (0.23) | (0.84) | (0.06) |
| 9 | -0.23** | -0.04 | -0.20** | -0.22** | -0.25** | -0.15* | -0.13 | -0.23*** | -0.12 | -0.20 | -0.20** | -0.05 | -0.17* | -0.08 | -0.18* | -0.05 | -0.12 | -0.16** | -0.22** | -0.09 |
| | (-2.37) | ( -0.25) | (-2.05) | (-2.52) | (-2.05) | (-1.65) | (-1.20) | (-2.56) | (-1.46) | (-1.68) | (-2.25) | (-0.52) | (-1.88) | (-0.75) | (-2.24) | (-0.37) | (-1.46) | (-2.02) | (-2.04) | (-0.83) |
| H | -0.28*** | -0.36*** | -0.21** | -0.22** | -0.18** | -0.25*** | -0.31*** | -0.24*** | -0.25** | -0.23** | -0.22*** | -0.24*** | -0.29*** | -0.34*** | -0.31*** | -0.39*** | -0.25** | -0.31*** | -0.27*** | -0.28*** |
| | (-2.89) | (-3.10) | (-2.35) | (-2.57) | (-2.25) | (-2.94) | (-4.03) | (-2.83) | (-2.35) | (-2.64) | (-2.73) | (-3.87) | (-2.95) | (-3.50) | (-3.71) | (-3.74) | (-2.35) | (-3.85) | (-3.52) | (-3.48) |
| H-L | -1.00*** | -0.99*** | -0.87*** | -0.92*** | -0.76*** | -0.90*** | -1.03*** | -0.92*** | -0.95*** | -0.87*** | -0.90*** | -0.80*** | -0.98*** | -1.07*** | -1.00*** | -1.06*** | -0.95*** | -1.03*** | -0.97*** | -0.86*** |
| | (-4.90) | (-4.20) | (-4.53) | (-4.38) | (-3.86) | (-4.72) | (-5.15) | (-4.59) | (-4.28) | (-4.63) | (-4.59) | (-4.58) | (-4.85) | (-4.95) | (-4.33) | (-4.43) | (-4.28) | (-5.23) | (-4.48) | (-4.12) |

Table 6: **Stock-level regressions:** This table reports estimates from regressing the next month's excess returns on Γ and a set of predictive variables using Fama and MacBeth (1973) regressions and panel regressions. Observations are both value-weighted (panel A and C) and equally-weighted (panel B and D). Regression specification (1) has no control variables. Regression specification (2) adds implied volatility (IV), Call Volume / Total option volume (Call Vol.), Call open interest / Total option open interest (Call OI), and a range of price-based control variables: market beta, total return volatility, idiosyncratic volatility, realized volatility, 1-year momentum, 1-month reversal, and the illiquidity measure of Amihud (2002). Specification (3) subsequently adds accounting control variables: book-to-market ratio, return on equity, investment/assets, operating profitability, and cash profitability. All coefficients are multiplied by 100. The constant is omitted for brevity. Both time - and firm fixed effects are included in the panel regressions. Newey-West t-statistics are reported between parentheses for Fama-Macbeth regressions. Two-way cluster (by stock and date) adjusted t-statistics are given in parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | **A: Value-weighted FMB** | | | **B: Equal-weighted FMB** | | | **C: Value-weighted Panel** | | | **D: Equal-weighted Panel** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) |
| Γ | -18.65*** | -18.08*** | -16.78*** | -17.97*** | -12.08*** | -12.37*** | -11.33*** | -10.58*** | -9.58*** | -12.53*** | -8.93*** | -8.18*** |
| | (-3.80) | (-3.21) | (-3.48) | (-3.15) | (-3.02) | (-3.21) | (-4.90) | (-3.60) | (-3.24) | (-5.64) | (-3.72) | (-3.49) |
| IV | | 0.33*** | 0.56*** | | 0.79*** | 0.90*** | | 2.43*** | 2.49*** | | 2.24*** | 2.59*** |
| | | (4.42) | (3.31) | | (6.59) | (6.03) | | (3.47) | (3.01) | | (8.68) | (7.96) |
| Call Vol. | | -0.08 | -0.12 | | 0.41** | 0.45** | | 0.07 | 0.05 | | 0.75*** | 0.73*** |
| | | (-0.34) | (-0.46) | | (2.15) | (2.45) | | (0.24) | (0.15) | | (5.25) | (5.22) |
| Call OI | | 0.15 | 0.14 | | 0.35** | 0.28 | | 0.69 | 0.42 | | 1.01*** | 0.79*** |
| | | (0.28) | (0.29) | | (2.03) | (1.62) | | (1.59) | (0.94) | | (5.19) | (4.02) |
| Obs. | 406K | 391K | 363K | 406K | 391K | 363K | 406K | 391K | 363K | 406K | 391K | 363K |
| $R^2$ | 1.78% | 16.35% | 19.17% | 0.43% | 10.01% | 11.49% | 0.10% | 0.46% | 0.61% | 0.02% | 0.97% | 1.15% |
| Price Controls | NO | YES | YES | NO | YES | YES | NO | YES | YES | NO | YES | YES |
| Acc. Controls | NO | NO | YES | NO | NO | YES | NO | NO | YES | NO | NO | YES |

Table 7: **Predicting extreme returns:** This table reports estimates from regressing the next month's 'extreme return' indicator on $\Gamma$ and a set of predictive variables using panel logit regressions. In panels A, B and C, the indicator variable takes value one when the next month's absolute return is larger than 25%, 50%, and 75%, else zero, respectively. Regression specification (1) has no control variables. Regression specification (2) adds call volume / total option volume (Call Vol.) and call open interest / total option open interest (Call OI), and a range of price-based control variables: 1-year momentum, 1-month reversal, and the illiquidity measure of Amihud (2002). Specification (3) subsequently adds accounting control variables: book-to-market ratio, return on equity, investment/assets, operating profitability, and cash profitability. Time fixed effects are included. One-way cluster (by date) adjusted t-statistics are given in parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | **A:** $I[abs(R_{t+1}) > 25\%]$ | | | **B:** $I[abs(R_{t+1}) > 50\%]$ | | | **C:** $I[abs(R_{t+1}) > 75\%]$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) |
| Gamma | -4.21*** | -11.55*** | -11.55*** | -4.03*** | -13.36*** | -13.14*** | -2.85** | -11.53*** | -12.63*** |
| | (-14.76) | (-15.72) | (-14.43) | (-5.51) | (-5.54) | (-4.91) | (-2.10) | (-2.32) | (-2.26) |
| MOM | | 0.12*** | 0.11*** | | 0.12*** | 0.11*** | | 0.08*** | 0.04 |
| | | (14.26) | (10.68) | | (8.68) | (6.54) | | (2.94) | (0.98) |
| SREV | | -0.95*** | -0.97*** | | -1.97*** | -1.99*** | | -2.89*** | -2.66*** |
| | | (-16.88) | (-15.64) | | (-13.60) | (-11.85) | | (-9.90) | (-7.78) |
| Call. Vol | | 0.33*** | 0.32*** | | 0.65*** | 0.70*** | | 1.03*** | 1.29*** |
| | | (7.30) | (6.53) | | (5.06) | (4.87) | | (3.77) | (4.14) |
| Call OI | | 0.12** | 0.07 | | 0.01 | -0.06 | | 0.11 | -0.11 |
| | | (2.25) | (1.11) | | (0.03) | (-0.36) | | (0.34) | (-0.30) |
| Obs. | 406K | 391K | 363K | 406K | 391K | 363K | 406K | 391K | 363K |
| Acc. Controls | NO | NO | YES | NO | NO | YES | NO | NO | YES |

Figure 3: **Gamma premium over time:** This figure shows the gamma premium over time. The solid line depicts the six-month moving average of the monthly slope coefficient of the net gamma exposure (Table 6 column 1). The grey-shaded area indicate periods in which the NBER recession indicator equals 1 (i.e. the economy is in a recession). The sample runs from February 1996 until December 2021.

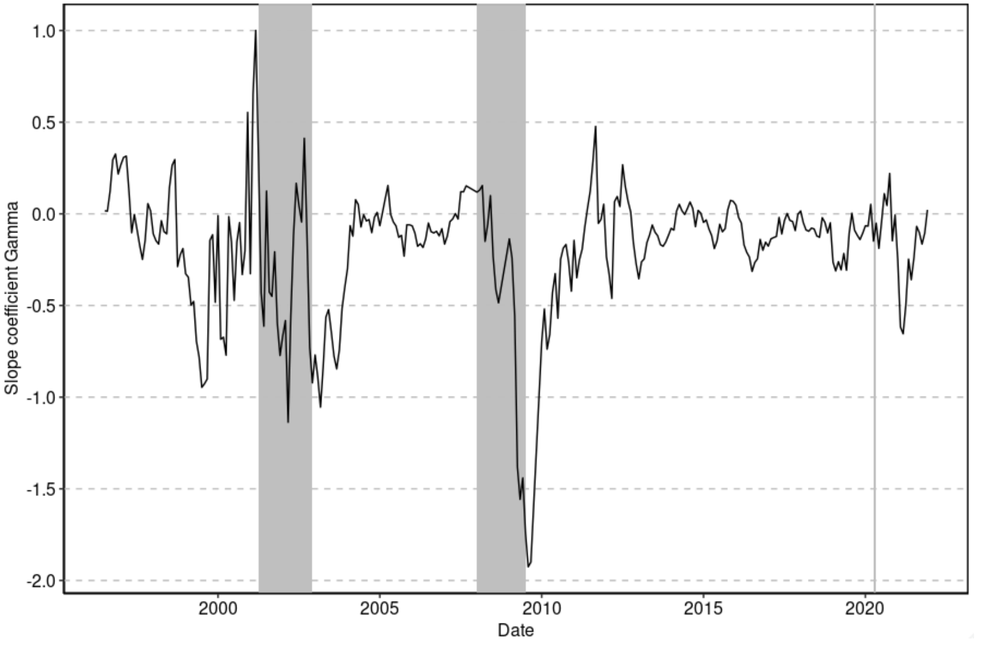


Table 8: **Time-varying gamma premium:** this table presents the estimates from regressing the estimated gamma premium (from table 6 column 1) on a set of macroeconomic indicators. CFNAI denotes the Chicago Fed National Activity Index. Sentiment denotes the sentiment measure of baker2006investor. FUNC is the financial uncertainty index of jurado2015measuring. All regressors are standardized by their full-sample mean and standard deviation. Newey-West t-statistics are reported between parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| CFNAI | $0.12^{**}$ | | | | $0.091^{***}$ |
| | (3.32) | | | | (6.51) |
| VIX | | $-0.05$ | | | $0.19^{*}$ |
| | | (-0.83) | | | (1.80) |
| Sentiment | | | $0.12^{**}$ | | $0.14^{***}$ |
| | | | (2.02) | | (3.26) |
| FUNC | | | | $-0.15^{**}$ | $-0.29^{**}$ |
| | | | | (-2.64) | (-2.46) |
| Obs. | 302 | 302 | 302 | 302 | 302 |
| $R^2$ | 1.3% | 0.2% | 1.2% | 2.0% | 5.5% |

Table 9: **Spanning regressions:** The table shows the estimated intercepts $\alpha$ (annualized in percentages), slopes, t-statistics for the intercepts $t(a)$, $R^2$, and residual standard errors $s(e)$ from spanning regressions of each of the factors of a model on the 2-by-3 gamma factor. The factor models are the three-factor model of Fama and French (1993), the five-factor model of Fama and French (2015), the cash-based five-factor model of Fama and French (2018), and Hou et al. (2015), all augmented with the momentum factor. The table also shows the maximum squared Sharpe ratio ($Sh^2(f)$) and the marginal contribution of the gamma factor to a model's $Sh^2(f)$, that is, $a^2/s^2(e)$. Panel A uses long-minus-short factors. Panel B (C) uses the long (short) legs in spanning regressions. The data runs from February 1996 until December 2021.

| **A: Long-Short** | $\hat{\alpha}$ | $Mkt$ | $SMB$ | $HML$ | $RMW$ | $CMA$ | $UMD$ | $IA$ | $ROE$ | $t(a)$ | $R^2$ | $s(e)$ | $Sh_f^2$ | $\hat{\alpha}^2/s^2(e)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FF3 | 3.11 | 0.066 | 0.164 | -0.183 | | | -0.262 | | | 2.260 | 0.414 | 0.018 | 0.073 | 0.021 |
| FF5 | 4.06 | 0.039 | 0.093 | -0.069 | -0.196 | -0.109 | -0.255 | | | 3.122 | 0.449 | 0.018 | 0.124 | 0.037 |
| FF5$_c$ | 4.09 | 0.032 | 0.102 | -0.098 | -0.203 | -0.094 | -0.252 | | | 3.096 | 0.444 | 0.018 | 0.131 | 0.037 |
| Q | 4.64 | 0.036 | 0.124 | | | | -0.222 | -0.130 | -0.216 | 3.209 | 0.416 | 0.018 | 0.238 | 0.046 |
| **B: Long** | $\hat{\alpha}$ | $Mkt$ | $SMB$ | $HML$ | $RMW$ | $CMA$ | $UMD$ | $IA$ | $ROE$ | $t(a)$ | $R^2$ | $s(e)$ | $Sh_f^2$ | $\hat{\alpha}^2/s^2(e)$ |
| FF3 | 2.18 | 0.787 | 0.781 | -0.157 | | | -0.351 | | | 2.506 | 0.956 | 0.011 | 0.079 | 0.026 |
| FF5 | 2.84 | 1.009 | 0.910 | -0.027 | -0.253 | -0.241 | -0.360 | | | 3.596 | 0.960 | 0.011 | 0.129 | 0.047 |
| FF5$_c$ | 2.83 | 0.984 | 0.893 | -0.031 | -0.220 | -0.226 | -0.359 | | | 3.532 | 0.959 | 0.011 | 0.135 | 0.046 |
| Q | 3.31 | 0.987 | 0.899 | | | | -0.342 | -0.204 | -0.297 | 3.776 | 0.958 | 0.011 | 0.286 | 0.061 |
| **C: Short** | $\hat{\alpha}$ | $Mkt$ | $SMB$ | $HML$ | $RMW$ | $CMA$ | $UMD$ | $IA$ | $ROE$ | $t(a)$ | $R^2$ | $s(e)$ | $Sh_f^2$ | $\hat{\alpha}^2/s^2(e)$ |
| FF3 | 0.20 | -1.069 | 1.609 | 0.456 | | | -0.031 | | | 0.208 | 0.921 | 0.013 | 0.040 | 0.000 |
| FF5 | 0.61 | -0.979 | 1.438 | 0.068 | 0.146 | 0.337 | -0.074 | | | 0.695 | 0.929 | 0.012 | 0.053 | 0.002 |
| FF5$_c$ | 0.67 | -0.791 | 1.228 | 0.056 | 0.252 | 0.273 | -0.090 | | | 0.777 | 0.932 | 0.012 | 0.050 | 0.002 |
| Q | 1.00 | -0.877 | 1.357 | | | | 0.431 | 0.103 | -0.085 | 1.144 | 0.927 | 0.012 | 0.111 | 0.005 |

Table 10: **Daily and weekly stock-level regressions:** This table reports estimates from regressing future excess returns on Γ and a set of predictive variables using Fama and MacBeth (1973) regressions, whereby observations are vale-weighted. We regress the next day and next week excess return on the net gamma exposure in panel (A) and (B), respectively. Regression specifications 5 and 6 add a range of price-based control variables: market beta, total return volatility, idiosyncratic volatility, 1-year momentum, 1-month reversal, and the illiquidity measure of Amihud (2002). All coefficients are multiplied by 100. The constant is omitted for brevity. Newey-West t-statistics are reported between parentheses for Fama-Macbeth regressions. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | **Panel A: Daily horizon** | | | | | | **Panel B: Weekly horizon** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Γ | -3.13*** | -3.06*** | -2.23*** | -2.59*** | -3.31*** | -2.91*** | -6.04*** | -5.62*** | -4.95*** | -5.25*** | -6.75*** | -6.39*** |
| | (-3.02) | (-2.97) | (-5.91) | (-7.58) | (-7.79) | (-7.78) | (-3.11) | (-2.97) | (-3.62) | (-4.67) | (-5.13) | (-5.53) |
| $R_{t-1}$ | | -0.45** | -0.75*** | -1.46*** | -1.50*** | -2.00*** | | -3.11*** | -3.59*** | -4.31*** | -4.27*** | -4.34*** |
| | | (-2.38) | (-3.80) | (-7.63) | (-7.80) | (-10.76) | | (-6.85) | (-7.80) | (-9.79) | (-9.70) | (-10.12) |
| $R_{t-1} \times \Gamma$ | | | 1.51 | 13.51 | 10.40 | 13.75* | | | 9.87 | 23.69 | 15.07 | 22.59 |
| | | | (0.16) | (1.62) | (1.31) | (1.81) | | | (0.47) | (1.28) | (0.95) | (1.51) |
| IV | | | | -0.10* | -0.10** | -0.42*** | | | | 0.05 | 0.06 | -0.10 |
| | | | | (-1.97) | (-1.85) | (-5.23) | | | | (0.23) | (0.24) | (-0.60) |
| Call Vol. | | | | | 0.07*** | 0.07*** | | | | | 0.08*** | 0.08*** |
| | | | | | (10.28) | (10.26) | | | | | (3.72) | (3.80) |
| Call OI. | | | | | 0.13*** | 0.08*** | | | | | 0.29*** | 0.27*** |
| | | | | | (5.54) | (4.12) | | | | | (2.82) | (3.23) |
| Obs. | 10.65M | 10.64M | 10.64M | 10.57M | 8.70M | 8.39M | 10,63M | 10,63M | 10,63M | 10.56M | 8.69M | 8.38M |
| $R^2$ | 1.61% | 3.78% | 4.40% | 8.57% | 9.59% | 16.39% | 1.69% | 3.70% | 4.25% | 8.62% | 9.70% | 16.74% |
| Controls | NO | NO | NO | NO | NO | YES | NO | NO | NO | NO | NO | YES |

Table 11: **Decomposing the net gamma exposure:** This table reports estimates from regressing monthly excess returns on Γ and a set of predictive variables using fama1973risk regressions and panel regressions. Observations are both value-weighted (panel A and C) and equally-weighted (panel B and D). The net gamma exposure is decomposed in two different ways. First, the net gamma exposure can be decomposed into an 'near-the-money' component ($\Gamma_{NTM}$), 'out-the-money' component ($\Gamma_{OTM}$), and 'in-the-money' component ($\Gamma_{ITM}$). An option is classified as "near-the-money" whenever the absolute values of the natural log of the ratio of the stock price to the exercise price less than 0.1. When the value exceeds 0.1, a call (put) option is "in-the-money" ("out-the-money"). Vice verse, when this value is below -0.1, a call (put) option is out-the-money (in-the-money). The second decomposition is in terms of option expiration: an option is considered as "fast" when it expires within the next month, else it is classified as "slow". This leads to $\Gamma_{fast}$ and $\Gamma_{slow}$ respectively. All coefficients are multiplied by 100. The constant is omitted for brevity. Both time - and firm fixed effects are included in panel regressions. Newey-West t-statistics are reported between parentheses for Fama-Macbeth regressions. Two-way cluster (by stock and date) adjusted t-statistics are given in parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | A: Value-weighted FMB | | | B: Equal-weighted FMB | | | C: Value-weighted Panel | | | D: Equal-weighted Panel | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) |
| Γ | -16.78*** | | | -12.37*** | | | -9.58*** | | | -8.18*** | | |
| | (-3.48) | | | (-3.21) | | | (-3.24) | | | (-3.49) | | |
| $\Gamma_{NTM}$ | | -13.25*** | | | -11.05*** | | | -11.06*** | | | -9.08*** | |
| | | (-4.58) | | | (-3.88) | | | (-3.43) | | | (-3.78) | |
| $\Gamma_{OTM}$ | | -29.17*** | | | -26.35*** | | | -47.20*** | | | -34.88*** | |
| | | (-3.62) | | | (-4.40) | | | (-3.67) | | | (-6.34) | |
| $\Gamma_{ITM}$ | | -34.83* | | | -1.59 | | | -33.37* | | | 2.62 | |
| | | (-1.91) | | | (-0.14) | | | (-1.89) | | | (0.24) | |
| $\Gamma_{fast}$ | | | 18.62 | | | 5.64 | | | -4.95 | | | -8.21** |
| | | | (0.60) | | | (0.34) | | | (-1.04) | | | (-2.54) |
| $\Gamma_{slow}$ | | | -26.25*** | | | -14.39*** | | | -13.13*** | | | -8.15*** |
| | | | (-3.40) | | | (-3.36) | | | (-3.25) | | | (-2.82) |
| IV | 1.32*** | 1.38*** | 1.13*** | 1.79*** | 1.82*** | 1.81*** | 3.72*** | 3.71*** | 3.70*** | 5.30*** | 5.26*** | 5.30*** |
| | (3.31) | (3.62) | (2.81) | (6.03) | (5.93) | (5.82) | (3.01) | (2.98) | (2.98) | (7.96) | (7.89) | (7.96) |
| Call Vol. | -0.12 | -0.15 | -0.18 | 0.45** | 0.47** | 0.43** | 0.05 | 0.06 | 0.01 | 0.73*** | 0.75*** | 0.73*** |
| | (-0.46) | (-0.55) | (-0.64) | (2.45) | (2.65) | (2.45) | (0.15) | (0.19) | (0.04) | (5.22) | (5.32) | (5.21) |
| Call OI | 0.14 | 0.56 | 0.31 | 0.28 | 0.39** | 0.33** | 0.42 | 1.36*** | 0.52 | 0.79*** | 1.04*** | 0.79*** |
| | (0.29) | (1.17) | (0.61) | (1.62) | (2.19) | (2.07) | (0.94) | (3.12) | (1.16) | (4.02) | (5.43) | (3.99) |
| Obs. | 363K | 363K | 363K | 363K | 363K | 363K | 363K | 363K | 363K | 363K | 363K | 363K |
| $R^2$ | 19.17% | 20.24% | 19.58% | 11.49% | 11.78% | 11.56% | 0.61% | 0.69% | 0.62% | 1.15% | 1.17% | 1.15% |
| Price Controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Acc. Controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |

Table 12: **Controlling for option-based predictors:** This table reports estimates from regressing the next month's excess returns on Γ and a set of predictive variables using Fama and MacBeth (1973) regressions, whereby observations are value-weighted. Regression specification (12) adds a range of price-based control variables: market beta, total return volatility, idiosyncratic volatility, realized volatility, 1-year momentum, 1-month reversal, and the illiquidity measure of Amihud (2002). Specification (13) subsequently adds accounting control variables: book-to-market ratio, return on equity, investment/assets, operating profitability, and cash profitability. All coefficients are multiplied by 100. The constant is omitted for brevity. Newey-West t-statistics are reported between parentheses for Fama-Macbeth regressions. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Γ | -18.65*** | -15.82*** | -16.33*** | -17.25*** | -17.63*** | -17.64*** | -18.94*** | -18.87*** | -18.81*** | -19.34*** | -21.27*** | -20.83*** | -18.99*** |
| | (-3.80) | (-3.36) | (-3.35) | (-3.62) | (-3.69) | (-2.82) | (-2.98) | (-2.99) | (-2.92) | (-3.18) | (-4.34) | (-3.66) | (-4.14) |
| RV-IV | | -1.28* | -0.94 | -0.82 | -0.63 | -0.60 | 0.19 | 0.12 | 0.20 | -0.06 | 1.30 | 1.18 | 1.13** |
| | | (-1.68) | (-1.12) | (-1.00) | (-0.73) | (-0.72) | (0.19) | (0.12) | (0.19) | (-0.05) | (1.13) | (1.49) | (1.96) |
| $IV_{skew}$ | | | 0.58 | 1.29* | 1.65* | 1.29** | 1.10* | 1.05* | 1.00* | 0.97* | 1.47*** | 1.33*** | 1.22*** |
| | | | (0.90) | (1.90) | (2.32) | (2.04) | (1.95) | (1.86) | (1.85) | (1.77) | (3.12) | (3.52) | (3.14) |
| VoV | | | | -1.73*** | -1.80*** | -1.56*** | -1.49*** | -1.37*** | -1.39*** | -1.30*** | -1.20*** | -0.72*** | -0.76*** |
| | | | | (-2.87) | (-3.08) | (-2.96) | (-2.84) | (-2.77) | (-2.85) | (-2.75) | (-2.90) | (-2.65) | (-2.68) |
| CPIV | | | | | 4.05** | 4.51** | 4.63** | 4.78** | 4.82** | 4.65** | 5.22** | 4.95*** | 5.16*** |
| | | | | | (2.33) | (2.54) | (2.56) | (2.55) | (2.62) | (2.42) | (2.57) | (3.19) | (3.11) |
| DOI | | | | | | -0.01 | 0.01 | 0.03 | 0.01 | -0.03 | 0.03 | 0.02 | 0.01 |
| | | | | | | (-0.14) | (0.11) | (0.29) | (0.08) | (-0.26) | (0.42) | (0.32) | (0.14) |
| IV | | | | | | | 0.71 | 0.69 | 0.71 | 0.52 | 0.40 | 1.33 | 1.09 |
| | | | | | | | (0.92) | (0.93) | (0.92) | (0.71) | (0.52) | (2.63) | (2.07) |
| Call Vol. | | | | | | | | -0.16 | -0.16 | -0.13 | -0.05 | -0.03 | -0.06 |
| | | | | | | | | (-0.71) | (-0.68) | (-0.52) | (-0.21) | (-0.11) | (-0.26) |
| Call OI. | | | | | | | | | 0.12 | 0.29 | 0.57** | 0.49* | 0.47* |
| | | | | | | | | | (0.25) | (0.99) | (2.19) | (1.82) | (1.84) |
| Δ | | | | | | | | | | 0.06 | -0.71*** | -0.70*** | -0.71*** |
| | | | | | | | | | | (0.88) | (-3.35) | (-3.20) | (-3.16) |
| O/S | | | | | | | | | | | 6.52*** | 6.59*** | 6.65*** |
| | | | | | | | | | | | (3.04) | (3.00) | (2.97) |
| Obs. | 406K | 393K | 361K | 361K | 354K | 354K | 354K | 353K | 353K | 353K | 353K | 349K | 324K |
| $R^2$ | 1.79% | 7.45% | 7.96% | 8.40% | 8.82% | 10.19% | 10.72% | 11.12% | 11.56% | 12.59% | 14.05% | 20.65% | 23.45% |
| Price Controls | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | YES |
| Acc. Controls | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | YES | YES |

Table 13: **Volatility and net gamma exposure:** This table reports estimates from regressing the next month's realized volatility on $\Gamma$ and a set of predictive variables using Fama and MacBeth (1973) regressions and panel regressions. Observations are both value-weighted (panel A and C) and equally-weighted (panel B and D). Regression specification (1) has no control variables. Regression specification (2) adds implied volatility (IV), Call Volume / Total option volume (Call Vol.), Call open interest / Total option open interest (Call OI), and a range of price-based control variables: market beta, total return volatility, idiosyncratic volatility, realized volatility, 1-year momentum, 1-month reversal, and the illiquidity measure of Amihud (2002). Specification (3) subsequently adds accounting control variables: book-to-market ratio, return on equity, investment/assets, operating profitability, and cash profitability. All coefficients are multiplied by 100. The constant is omitted for brevity. Both time - and firm fixed effects are included in panel regressions. Newey-West t-statistics are reported between parentheses for Fama-Macbeth regressions. Two-way cluster (by stock and date) adjusted t-statistics are given in parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | **A: Value-weighted FMB** | | | **B: Equal-weighted FMB** | | | **C: Value-weighted Panel** | | | **D: Equal-weighted Panel** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) |
| Gamma | -12.92*** | -3.41*** | -3.24*** | -12.19*** | -3.29*** | -3.20*** | -1.87*** | -1.55*** | -1.51*** | -4.00*** | -2.54*** | -2.51*** |
| | (-3.58) | (-4.91) | (-4.60) | (-3.51) | (-3.92) | (-4.09) | (-6.40) | (-7.11) | (-6.53) | (-13.64) | (-12.46) | (-12.33) |
| IV | | 7.07*** | -4.41 | | 7.69*** | 1.60 | | 3.26*** | 3.23*** | | 2.53*** | 2.44*** |
| | | (7.78) | (-1.01) | | (11.54) | (0.78) | | (22.10) | (21.04) | | (34.65) | (32.76) |
| Call Vol. | | 3.52*** | 7.01*** | | 3.06*** | 7.67*** | | -0.16*** | -0.16*** | | -0.06*** | -0.06*** |
| | | (29.62) | (7.65) | | (30.98) | (11.27) | | (-5.78) | (-5.66) | | (-4.51) | (-4.51) |
| Call OI | | -0.07* | -0.05 | | -0.04* | -0.03 | | 0.54*** | 0.54*** | | 0.25*** | 0.24*** |
| | | (-1.77) | (-1.64) | | (-1.89) | (-1.47) | | (15.43) | (14.86) | | (13.48) | (13.01) |
| Obs. | 406K | 391K | 363K | 406K | 391K | 363K | 406K | 391K | 363K | 406K | 391K | 363K |
| $R^2$ | 5.38% | 51.41% | 52.18% | 1.31% | 43.11% | 43.48% | 0.32% | 21.21% | 21.11% | 0.26% | 15.39% | 15.24% |
| Price Controls | NO | YES | YES | NO | YES | YES | NO | YES | YES | NO | YES | YES |
| Acc. Controls | NO | YES | YES | NO | YES | YES | NO | YES | YES | NO | YES | YES |

Table 14: **Information Gamma and Hedge Gamma:** This table reports estimates from regressing next month's realized volatility on $\Gamma$ and a set of predictive variables using Fama and MacBeth (1973) regressions and panel regressions. Observations are both value-weighted (panel A and C) and equally-weighted (panel B and D). $\Gamma(t-\tau, S_t)$ is the net gamma exposure using the open interest at $t-\tau$. $\Gamma_{info}$ is the information gamma, defined as the difference between $\Gamma$ and $\Gamma(t-\tau, S_t)$. $\Gamma(t-\tau, S_{t-\tau})$ is the net gamma exposure at time $t-\tau$. $\Gamma_{hedge}$ is the hedge gamma, defined as the difference between $\Gamma_{info}$ and $\Gamma(t-\tau, S_{t-\tau})$. All coefficients are multiplied by 100. The constant is omitted for brevity. Both time - and firm fixed effects are included in panel regressions. Newey-West t-statistics are reported between parentheses for Fama-Macbeth regressions. Two-way cluster (by stock and date) adjusted t-statistics are given in parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | **A: Value-weighted FMB** | | | **B: Equal-weighted FMB** | | | **C: Value-weighted Panel** | | | **D: Equal-weighted Panel** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) |
| $\Gamma$ | -3.24*** | | | -3.20*** | | | -1.51*** | | | -2.51*** | | |
| | (-4.60) | | | (-4.09) | | | (-6.53) | | | (-12.33) | | |
| $\Gamma(t-\tau, S_t)$ | | -3.25*** | | | -3.44*** | | | -1.78*** | | | -2.90*** | |
| | | (-4.57) | | | (-3.82) | | | (-7.00) | | | (-13.44) | |
| $\Gamma_{info}$ | | 0.45 | -0.03 | | 2.18 | 1.43 | | 0.69 | 0.94* | | -0.05 | 0.23 |
| | | (0.22) | (-0.02) | | (1.27) | (1.62) | | (1.14) | (1.84) | | (-0.14) | (0.64) |
| $\Gamma(t-\tau, S_{t-\tau})$ | | | -3.00*** | | | -2.93*** | | | -1.61*** | | | -2.59*** |
| | | | (-5.00) | | | (-3.80) | | | (-6.11) | | | (-11.22) |
| $\Gamma_{hedge}$ | | | -4.13** | | | -6.81*** | | | -2.38*** | | | -4.93*** |
| | | | (-2.60) | | | (-3.05) | | | (-4.14) | | | (-10.39) |
| IV | 3.44*** | 3.79*** | 3.52*** | 2.94*** | 3.34*** | 3.08*** | 3.23*** | 3.34*** | 3.3*** | 2.44*** | 2.63*** | 2.63*** |
| | (27.53) | (14.99) | (27.03) | (30.43) | (14.09) | (33.36) | (21.04) | (20.90) | (20.90) | (32.76) | (34.63) | (34.62) |
| Call Vol. | -0.05 | -0.06 | -0.08** | -0.03 | -0.02 | -0.04* | -0.16*** | -0.18*** | -0.18*** | -0.06*** | -0.06*** | -0.06*** |
| | (-1.64) | (-1.57) | (-2.18) | (-1.47) | (-0.69) | (-1.91) | (-5.66) | (-5.78) | (-6.23) | (-4.51) | (-4.64) | (-4.83) |
| Call OI. | 0.51*** | 0.54*** | 0.51*** | 0.26*** | 0.29*** | 0.26*** | 0.54*** | 0.57*** | 0.56*** | 0.24*** | 0.26*** | 0.26*** |
| | (15.42) | (12.65) | (14.43) | (12.11) | (9.76) | (9.49) | (14.86) | (15.34) | (14.58) | (13.01) | (13.61) | (13.17) |
| Obs. | 363K | 351K | 351K | 363K | 351K | 351K | 363K | 351K | 351K | 363K | 351K | 351K |
| $R^2$ | 52.18% | 54.36% | 54.64% | 43.48% | 45.76% | 45.88% | 21.11% | 21.60% | 21.65% | 15.24$ | 15.94% | 15.95% |
| Price Controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Acc. Controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |

Table 15: **Earnings announcement returns and net gamma exposure:** This table reports estimates from regression next day's excess return on Γ, an earnings announcement dummy, the interaction between Γ and the dummy, and a set of predictive control variables using Fama and MacBeth (1973) regressions. In columns 1-3, the earnings announcement dummy takes value 1 (else 0) on day $t$ if there is an earnings announcement. In columns 4-5, the earnings announcement dummy takes value 1 (else 0) on days $[t-1, t+2]$ if there is an earnings announcement on day $t$. Observations are value-weighted. All coefficients are multiplied by 100. The constant is omitted for brevity. Newey-West t-statistics are reported between parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Γ | -3.13*** | -2.01*** | -2.68*** | -1.82*** | -2.70*** |
| | (-3.02) | (-5.07) | (-6.92) | (-5.27) | (-6.67) |
| I[Earnings] | | 5.22 | 0.16 | 3.85 | 2.58 |
| | | (1.05) | (0.96) | (1.02) | (1.04) |
| Γ×I[Earnings] | | -1424.32 | 18.82 | -1119.08 | -741.88 |
| | | (-0.97) | (0.40) | (-1.00) | (-1.01) |
| $R_{t-1}$ | | | -1.76*** | | -1.91*** |
| | | | (-8.56) | | (-10.14) |
| Γ×$R_{t-1}$ | | | 19.04** | | 13.12 |
| | | | (2.07) | | (1.64) |
| Obs. | 10.65M | 10.64M | 8.39M | 10.64M | 8.39M |
| $R^2$ | 1.61% | 2.74% | 17.46% | 2.99% | 17.63% |
| Controls | NO | NO | YES | NO | YES |

Table 16: **Predicting future trading volume:** This table reports estimates from regressing the next month's percentage change in trading volume on the absolute Γ and a set of predictive variables using Fama and MacBeth (1973) regressions and panel regressions. Observations are both value-weighted (panel A and C) and equally-weighted (panel B and D). Regression specification (1) has no control variables. Regression specification (2) adds implied volatility (IV), Call Volume / Total option volume (Call Vol.), Call open interest / Total option open interest (Call OI), and a range of price-based control variables: market beta, total return volatility, idiosyncratic volatility, realized volatility, 1-year momentum, 1-month reversal, and the illiquidity measure of Amihud (2002). Specification (3) subsequently adds accounting control variables: book-to-market ratio, return on equity, investment/assets, operating profitability, and cash profitability. The constant is omitted for brevity. Both time - and firm fixed effects are included in the panel regressions. Newey-West t-statistics are reported between parentheses for Fama-Macbeth regressions. Two-way cluster (by stock and date) adjusted t-statistics are given in parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | A: Value-weighted FMB | | | B: Equal-weighted FMB | | | C: Value-weighted Panel | | | D: Equal-weighted Panel | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) |
| $\|Gamma\|$ | 1.01*** | 1.06*** | 1.04*** | 3.04*** | 3.05*** | 2.99*** | 1.24*** | 1.01*** | 0.94*** | 3.18*** | 2.61*** | 2.51*** |
| | (5.69) | (4.90) | (4.83) | (5.17) | (6.43) | (6.40) | (7.79) | (4.41) | (4.13) | (23.92) | (20.94) | (20.59) |
| IV | | 0.72*** | 0.69*** | | 0.55*** | 0.53*** | | 0.80*** | 0.75*** | | 0.53*** | 0.51*** |
| | | (12.80) | (13.13) | | (10.41) | (9.79) | | (9.52) | (9.97) | | (14.59) | (13.45) |
| Call. Vol. | | -0.04*** | -0.04*** | | -0.00 | -0.00 | | -0.07*** | -0.06*** | | -0.00 | -0.00 |
| | | (-2.78) | (-3.05) | | (-0.40) | (-0.13) | | (-5.08) | (-5.11) | | (-0.32) | (-0.38) |
| Call OI | | 0.07*** | 0.07*** | | -0.00 | -0.01 | | 0.06*** | 0.06*** | | 0.02** | 0.01* |
| | | (4.03) | (4.24) | | (-0.17) | (-0.48) | | (3.71) | (3.55) | | (2.08) | (1.80) |
| Obs. | 406K | 391K | 363K | 406K | 391K | 363K | 406K | 391K | 363K | 406K | 391K | 363K |
| $R^2$ | 1.11% | 12.21% | 14.25% | 0.54% | 10.03% | 10.79% | 0.33% | 5.51% | 5.57% | 0.40% | 4.35% | 4.82% |
| Price Controls | NO | YES | YES | NO | YES | YES | NO | YES | YES | NO | YES | YES |
| Acc. Controls | NO | NO | YES | NO | NO | YES | NO | NO | YES | NO | NO | YES |

Table 17: **Dynamic return-volume relationship and gamma exposures:** This table reports estimates from regression the excess return at $t+1$ ($r_{t+1}$) on $r_t$, the interaction between $r_t$ and turnover, and the interaction between $r_t$, turnover and absolute net gamma exposures. Observations are value-weighted. All coefficients are multiplied by 100. The constant is omitted for brevity. Newey-West t-statistics are reported between parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $R_t$ | 3.43*** | 3.33*** | 1.20** | 1.25*** | 1.10*** |
| | (7.03) | (6.33) | (2.52) | (2.69) | (2.40) |
| $V_t \times R_t$ | 0.89*** | 1.00*** | 0.71*** | 0.75*** | 0.79*** |
| | (7.84) | (7.82) | (6.06) | (6.74) | (7.19) |
| $V_t \times R_t \times \|\|Gamma\|\|$ | | -16.31 | -6.95*** | -5.93*** | -6.55*** |
| | | (-1.05) | (-3.78) | (-3.31) | (-3.81) |
| IV | | | -0.04 | -0.33*** | -0.38*** |
| | | | (-0.76) | (-4.40) | (-4.74) |
| Call Vol. | | | 0.07*** | 0.06*** | 0.06*** |
| | | | (8.51) | (8.78) | (9.10) |
| Call OI | | | 0.03 | -0.00 | -0.02 |
| | | | (1.30) | (-0.10) | (-1.30) |
| $\beta_{MKT}$ | | | | 0.06*** | 0.05*** |
| | | | | (4.14) | (3.60) |
| VOL | | | | 24.37*** | 24.79*** |
| | | | | (2.85) | (2.74) |
| IVOL | | | | 3.48*** | 3.69*** |
| | | | | (5.02) | (5.38) |
| MOM | | | | | 0.01 |
| | | | | | (0.40) |
| SREV | | | | | -0.22*** |
| | | | | | (-6.09) |
| ILQ | | | | | 6.08*** |
| | | | | | (7.32) |
| Obs | 10.64M | 9.52M | 7.82M | 7.82M | 7.81M |
| $R^2$ | 3.27% | 4.25% | 9.82% | 14.11% | 16.85% |

## 4.9 Additional tables & figures

Table A.1: **Average portfolio characteristics:** This table reports the average characteristic of decile portfolios formed on the basis of the net gamma exposure, which measures the total outstanding gamma divided by the average daily dollar trading volume. At the end of month $t$ we sort stocks into ten portfolios based on their net gamma exposure, and hold this portfolio during month $t+1$. We compute the value-weighted average of a characteristic for each decile. The row labeled "L-H" is the self-financing high-minus-low portfolio, which reports the difference in the average characteristic value between portfolio H and portfolio L. The sample consists of stocks listed on NYSE/AMEX/NASDAQ for the period between January 1996 and December 2021 with share code 10 or 11. Newey-West t-statistics are reported between parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | MKT | ME | BM | $BM_m$ | RMW | CP | IA | NSI | CSI | ROE | MOM | SREV | VOL | IVOL | RV | ILQ | MAX | IV | CVOL | COI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| L | 1.01*** | 9.98*** | 0.38*** | 0.38*** | 0.86*** | 0.80** | 0.17*** | 0.01*** | 0.39** | 0.07*** | 0.14*** | -0.03*** | 0.00*** | 0.02*** | 0.02*** | 0.00*** | 0.39*** | 0.37*** | 0.48*** | 0.46*** |
| | (64.11) | (80.17) | (24.48) | (25.94) | (2.68) | (2.50) | (8.44) | (2.64) | (4.34) | (10.39) | (5.18) | (-7.83) | (5.43) | (12.42) | (15.83) | (3.81) | (4.34) | (20.19) | (83.53) | (81.58) |
| 2 | 1.06*** | 9.25*** | 0.43*** | 0.43*** | 0.91** | 0.85** | 0.22*** | 0.02*** | 0.42*** | 0.05*** | 0.20*** | -0.01*** | 0.00*** | 0.02*** | 0.02*** | 0.0*** 0 | 0.42*** | 0.41*** | 0.53*** | 0.51*** |
| | (67.32) | (82.59) | (24.01) | (23.30) | (2.06) | (1.98) | (5.24) | (4.08) | (5.25) | (6.44) | (5.03) | (-3.49) | (4.41) | (9.98) | (13.23) | (4.53) | (5.25) | (16.57) | (51.16) | (43.58) |
| 3 | 1.09*** | 8.88*** | 0.47*** | 0.45*** | 0.36*** | 0.34*** | 0.20*** | 0.03*** | 0.45*** | 0.05*** | 0.21*** | -0.00 | 0.00*** | 0.02*** | 0.02*** | 0.00*** | 0.45*** | 0.41*** | 0.59*** | 0.57*** |
| | (54.19) | (95.59) | (21.45) | (19.74) | (7.97) | (12.49) | (5.97) | (5.88) | (5.78) | (10.75) | (5.23) | (-1.13) | (4.93) | (9.65) | (12.62) | (7.26) | (5.78) | (17.09) | (58.78) | (52.32) |
| 4 | 1.10*** | 9.05*** | 0.46*** | 0.43*** | 0.33*** | 0.32*** | 0.20*** | 0.02*** | 0.47*** | 0.04*** | 0.22*** | 0.00 | 0.00*** | 0.02*** | 0.02 *** | 0.00*** | 0.47*** | 0.41*** | 0.61*** | 0.59*** |
| | (53.06) | (114.12) | (23.30) | (20.96) | (38.34) | (27.29) | (6.19) | (5.21) | (5.49) | (13.34) | (4.48) | (0.21) | (4.16) | (9.38) | (12.32) | (7.32) | (5.49) | (15.86) | (69.72) | (60.41) |
| 5 | 1.09*** | 9.27*** | 0.44*** | 0.41*** | 0.35*** | 0.36*** | 0.20*** | 0.02*** | 0.46*** | 0.06*** | 0.24*** | 0.01*** | 0.00*** | 0.02*** | 0.02*** | 0.00*** | 0.46*** | 0.40*** | 0.63*** | 0.60*** |
| | (79.94) | (148.52) | (23.75) | (18.55) | (22.63) | (21.05) | (7.39) | (5.56) | (5.37) | (4.04) | (4.95) | (2.71) | (4.30) | (9.87) | (14.12) | (5.09) | (5.37) | (17.14) | (76.55) | (63.16) |
| 6 | 1.10*** | 9.57*** | 0.43*** | 0.39*** | 0.41*** | 0.39*** | 0.20*** | 0.02*** | 0.47*** | 0.05*** | 0.24*** | 0.01*** | 0.00*** | 0.02*** | 0.02*** | 0.00*** | 0.47*** | 0.39*** | 0.63*** | 0.60*** |
| | (90.95) | (133.21) | (20.02) | (18.03) | (9.75) | (11.86) | (9.12) | (5.22) | (5.35) | (8.94) | (5.46) | (4.06) | (4.94) | (10.50) | (15.25) | (5.78) | (5.35) | (20.06) | (80.43) | (61.04) |
| 7 | 1.09*** | 9.80*** | 0.41*** | 0.38*** | 0.38*** | 0.36*** | 0.18*** | 0.02*** | 0.48*** | 0.06*** | 0.25*** | 0.02*** | 0.00*** | 0.02*** | 0.02*** | 0.00*** | 0.48*** | 0.38*** | 0.64*** | 0.60*** |
| | (99.25) | (157.30) | (18.26) | (17.12) | (27.59) | (15.20) | (10.53) | (4.75) | (4.96) | (11.21) | (5.77) | (6.74) | (5.18) | (11.67) | (14.89) | (5.28) | (4.96) | (21.40) | (81.35) | (59.01) |
| 8 | 1.06*** | 10.15*** | 0.40*** | 0.36*** | 0.75** | 0.72** | 0.17*** | 0.02*** | 0.47*** | 0.06*** | 0.23*** | 0.02*** | 0.00*** | 0.02*** | 0.02*** | 0.00*** | 0.47*** | 0.37*** | 0.64*** | 0.60*** |
| | (103.73) | (177.55) | (19.41) | (18.19) | (2.71) | (2.55) | (11.82) | (4.12) | (4.69) | (7.30) | (6.68) | (10.18) | (4.47) | (11.14) | (16.45) | (5.09) | (4.69) | (22.11) | (73.99) | (52.17) |
| 9 | 1.02*** | 10.62*** | 0.37*** | 0.33*** | 0.54*** | 0.50*** | 0.16*** | 0.01*** | 0.47*** | 0.07*** | 0.23*** | 0.03*** | 0.00*** | 0.02*** | 0.02*** | 0.00*** | 0.47*** | 0.35*** | 0.64*** | 0.59*** |
| | (66.87) | (151.27) | (17.29) | (17.53) | (5.56) | (4.95) | (12.00) | (3.30) | (4.56) | (10.01) | (6.18) | (12.85) | (5.98) | (12.17) | (16.24) | (5.06) | (4.56) | (21.95) | (76.96) | (54.86) |
| H | 0.90*** | 11.15*** | 0.35*** | 0.31*** | 0.60*** | 0.57*** | 0.14*** | 0.01** | 0.41*** | 0.07*** | 0.20*** | 0.04*** | 0.00*** | 0.01*** | 0.02*** | 0.00*** | 0.41*** | 0.32*** | 0.65*** | 0.60*** |
| | (30.11) | (81.61) | (18.44) | (18.54) | (4.98) | (4.60) | (16.17) | (2.22) | (4.33) | (14.47) | (8.79) | (17.86) | (4.15) | (8.27) | (16.71) | (3.51) | (4.33) | (23.53) | (72.70) | (51.85) |
| H-L | -0.10*** | 1.17*** | -0.03** | -0.07*** | -0.26 | -0.23 | -0.03* | -0.00 | 0.03 | -0.00 | 0.06*** | 0.06*** | -0.00*** | -0.00*** | -0.00*** | -0.00*** | 0.03 | -0.05*** | 0.1***7 | 0.14*** |
| | (-3.13) | (15.39) | (-2.25) | (-5.45) | (-1.13) | (-1.00) | (-1.84) | (-0.82) | (1.06) | (-0.29) | (3.77) | (15.18) | (-4.88) | (-10.46) | (-8.97) | (-3.73) | (1.06) | (-7.29) | (26.26) | (18.72) |

Table A.2: **Various sub-samples:** This table reports estimates from regressing monthly excess returns on $\Gamma$ and a set of predictive variables using panel regressions, whereby observations are weighted by their 1-month lagged market capitalization. Panel A uses the sample consisting of the largest 1000 firms in terms of market capitalization. Panel B uses the sample consisting of the 1000 most liquid firms according to the illiquidity measure of Amihud (2002). Panel C considers the sample of the top 1000 firms with most option trading volume in a month. Regression specification (1) has no control variables. Regression specification (2) adds market beta, total return volatility, idiosyncratic volatility, realized volatility, 1-year momentum, 1-month reversal, and the illiquidity measure of Amihud (2002) as control variable. Regression specification (3) subsequently adds implied volatility (IV), Call Volume / Total option volume (Call Vol.), Call open interest / Total option open interest (Call OI). Specification (4) adds accounting control variables: book-to-market ratio, return on equity, investment/assets, operating profitability, and cash profitability. All coefficients are multiplied by 100. The constant is omitted for brevity. Both time - and firm fixed effects are included in panel regressions. Newey-West t-statistics are reported between parentheses for Fama-Macbeth regressions. Two-way cluster (by stock and date) adjusted t-statistics are given in parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | **A: Top 1000 largest:** | | | | **B: Top 1000 most liquid:** | | | | **C: Top 1000 most option trading:** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) | (1) | ( 2) | (3) | (4) |
| Gamma | -11.36*** | -9.89*** | -10.84*** | -9.78*** | -11.41*** | -10.01*** | -10.91*** | -9.90*** | -11.54*** | -9.82*** | -11.14*** | -10.12*** |
| | (-4.94) | (-3.68) | (-3.67) | (-3.28) | (-4.97) | (-3.73) | (-3.69) | (-3.32) | (-4.92) | (-3.58) | (-3.60) | (-3.26) |
| IV | | | 3.78*** | 3.49** | | | 3.88*** | 3.56** | | | 4.56*** | 4.19*** |
| | | | (2.89) | (2.56) | | | (2.93) | (2.58) | | | (3.18) | (2.78) |
| Call Vol. | | | 0.02 | -0.00 | | | 0.01 | -0.01 | | | -0.03 | -0.02 |
| | | | (0.06) | (-0.01) | | | (0.03) | (-0.02) | | | (-0.06) | (-0.04) |
| Call OI | | | 0.66 | 0.38 | | | 0.64 | 0.37 | | | 0.99 | 0.64 |
| | | | (1.39) | (0.78) | | | (1.32) | (0.75) | | | (1.63) | (1.03) |
| Obs. | 301K | 297K | 296K | 277K | 301K | 301K | 300k | 281K | 302K | 294K | 294K | 272K |
| $R^2$ | 0.11% | 0.34% | 0.42% | 0.57% | 0.11% | 0.40% | 0.49% | 0.63% | 0.11% | 0.37% | 0.48% | 0.62% |
| Price Controls | NO | YES | YES | YES | NO | YES | YES | YES | NO | YES | YES | YES |
| Acc. Controls | NO | NO | NO | YES | NO | NO | NO | YES | NO | NO | NO | YES |

Table A.3: **Stock-level regressions with microcaps and without price filters:** This table reports estimates from regressing the next month's excess returns on the net gamma exposure and a set of predictive variables using Fama and MacBeth (1973) regressions and panel regressions. Observations are both value-weighted (panel A and C) and equally-weighted (panel B and D). Regression specification (1) has no control variables. Regression specification (2) adds implied volatility (IV), Call Volume / Total option volume (Call Vol.), Call open interest / Total option open interest (Call OI), and a range of price-based control variables: market beta, total return volatility, idiosyncratic volatility, realized volatility, 1-year momentum, 1-month reversal, and the illiquidity measure of Amihud (2002). Specification (3) subsequently adds accounting control variables: book-to-market ratio, return on equity, investment/assets, operating profitability, and cash profitability. All coefficients are multiplied by 100. The constant is omitted for brevity. Both time - and firm fixed effects are included in panel regressions. Newey-West t-statistics are reported between parentheses for Fama-Macbeth regressions. Two-way cluster (by stock and date) adjusted t-statistics are given in parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | **A: Value-weighted FMB** | | | **B: Equal-weighted FMB** | | | **C: Value-weighted Panel** | | | **D: Equal-weighted Panel** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) |
| Gamma | -17.41*** | -19.75*** | -18.46*** | -18.79*** | -18.92*** | -19.17*** | -10.23*** | -9.27*** | -8.16*** | -15.72*** | -15.29*** | -13.75*** |
| | (-3.52) | (-3.01) | (-3.19) | (-3.10) | (-3.90) | (-3.75) | (-4.63) | (-3.24) | (-2.84) | (-6.14) | (-5.13) | (-4.86) |
| IV | | 1.27*** | 1.12*** | | 0.89*** | 0.97*** | | 4.00*** | 3.66*** | | 5.03*** | 4.55*** |
| | | (3.46) | (3.09) | | (3.30) | (3.78) | | (3.63) | (3.21) | | (7.78) | (7.29) |
| Call Vol. | | 0.01 | -0.03 | | 0.80*** | 0.78*** | | 0.27 | 0.24 | | 1.26*** | 1.26*** |
| | | (0.04) | (-0.13) | | (5.54) | (5.98) | | (0.92) | (0.79) | | (8.23) | (8.26) |
| Call OI | | 0.05 | 0.09 | | 0.29** | 0.28* | | 0.50 | 0.22 | | 1.33*** | 1.07*** |
| | | (0.09) | (0.18) | | (2.23) | (1.97) | | (1.11) | (0.47) | | (6.14) | (5.19) |
| Obs. | 564K | 530K | 485K | 564K | 530K | 485K | 564K | 530K | 485K | 564K | 530K | 485K |
| $R^2$ | 1.71% | 15.23% | 17.93% | 0.32% | 8.16% | 9.29% | 0.07% | 0.44% | 0.57% | 0.02% | 0.88% | 0.93% |
| Price Controls | NO | YES | YES | NO | YES | YES | NO | YES | YES | NO | YES | YES |
| Acc. Controls | NO | NO | YES | NO | NO | YES | NO | NO | YES | NO | NO | YES |

Table A.4: **Sorting on alternative gamma definitions:** This table reports the performance of decile portfolios formed on the basis of Γ. In panel A: Γ is measured with a 1-day implementation lag. In panel B: Γ is measured as the average net gamma exposure within month $t$. At the end of month $t$ we sort stocks into ten portfolios based on their Γ, and hold this portfolio during month $t+1$. The results are shown for value-weighted portfolios whereby the breakpoints are based on the NYSE universe. Stocks with prices above $5 and classified as microcaps as of the portfolio formation are excluded. We report the time-series average of the net gamma exposure (Γ), the return ("R") in percentages, the Fama-French-Carhart four-factor alpha ("$\alpha_{3FM}$"), the Fama-French-Carhart five-factor alpha ("$\alpha_{5F}$"), the Fama-French-Carhart six-factor alpha ("$\alpha_{5FM}$"), Hou, Xue, and Zhang's extended q-factor model alpha ("$\alpha_{5Q}$"), and augmented with momentum ("$\alpha_{5QM}$") for each portfolio. The row labeled "L-H" is the self-financing high-minus-low portfolio, which reports the difference in between portfolio H and portfolio L. The sample consists of stocks listed on NYSE/AMEX/NASDAQ for the period between February 1996 and December 2021 with share code 10 or 11. Newey-West t-statistics are reported between parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | **Panel A: 1-day implementation lag:** | | | | | | | **Panel B: Average monthly gamma:** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Γ | $R$ | $\alpha_{3FM}$ | $\alpha_{5F}$ | $\alpha_{5FM}$ | $\alpha_{Q5}$ | $\alpha_{Q5M}$ | Γ | $R$ | $\alpha_{3FM}$ | $\alpha_{5F}$ | $\alpha_{5FM}$ | $\alpha_{Q5}$ | $\alpha_{Q5M}$ |
| L | -0.01*** | 1.36*** | 0.53*** | 0.48*** | 0.57*** | 0.51** | 0.51*** | -0.01*** | 1.23*** | 0.40*** | 0.24* | 0.34** | 0.38** | 0.38** |
| | (-11.79) | (5.24) | (3.37) | (3.11) | (3.72) | (2.47) | (2.94) | (-9.86) | (4.64) | (2.89) | (1.73) | (2.48) | (2.09) | (2.59) |
| 2 | -0.00*** | 1.06*** | 0.19* | 0.13 | 0.21** | 0.28** | 0.28** | -0.00*** | 1.02*** | 0.14 | 0.15 | 0.22* | 0.29** | 0.29** |
| | (-5.75) | (3.71) | (1.84) | (1.10) | (2.07) | (2.22) | (2.40) | (-4.59) | (3.61) | (1.22) | (1.10) | (1.91) | (2.19) | (2.43) |
| 3 | 0.00*** | 1.06*** | 0.14 | 0.10 | 0.18 | 0.23* | 0.23* | 0.00*** | 0.98*** | 0.07 | 0.08 | 0.15 | 0.19 | 0.19 |
| | (4.51) | (3.55) | (1.16) | (0.74) | (1.50) | (1.85) | (1.97) | (5.22) | (3.22) | (0.65) | (0.61) | (1.47) | (1.46) | (1.65) |
| 4 | 0.00*** | 1.19*** | 0.30** | 0.32*** | 0.37*** | 0.43*** | 0.43*** | 0.00*** | 1.04*** | 0.18 | 0.22* | 0.29** | 0.35*** | 0.35*** |
| | (10.18) | (4.93) | (2.45) | (2.66) | (3.01) | (2.74) | (2.90) | (11.55) | (3.26) | (1.30) | (1.87) | (2.64) | (2.94) | (3.07) |
| 5 | 0.00*** | 0.98*** | 0.03 | 0.11 | 0.13 | 0.13 | 0.13 | 0.00*** | 1.12*** | 0.23* | 0.35** | 0.38*** | 0.33*** | 0.34** |
| | (13.67) | (3.22) | (0.27) | (0.93) | (1.06) | (1.09) | (1.07) | (15.88) | (3.51) | (1.70) | (2.48) | (2.67) | (2.69) | (2.63) |
| 6 | 0.01*** | 0.95*** | 0.00 | 0.02 | 0.03 | -0.02 | -0.02 | 0.01*** | 1.09*** | 0.10 | 0.13 | 0.12 | 0.05 | 0.05 |
| | (16.47) | (3.38) | (0.03) | (0.30) | (0.34) | (-0.17) | (-0.17) | (18.85) | (4.20) | (0.82) | (1.08) | (1.09) | (0.49) | (0.50) |
| 7 | 0.01*** | 1.04*** | 0.17 | 0.14 | 0.16 | 0.15 | 0.15 | 0.01*** | 0.91*** | 0.01 | 0.05 | 0.06 | 0.12 | 0.12 |
| | (18.46) | (3.77) | (1.49) | (1.39) | (1.56) | (1.49) | (1.48) | (20.03) | (3.08) | (0.06) | (0.43) | (0.53) | (1.01) | (1.03) |
| 8 | 0.01*** | 0.85*** | -0.03 | 0.05 | 0.02 | -0.01 | -0.01 | 0.01*** | 0.87*** | 0.00 | -0.01 | -0.00 | 0.05 | 0.05 |
| | (19.46) | (3.04) | (-0.27) | (0.49) | (0.21) | (-0.08) | (-0.09) | (19.89) | (3.18) | (0.03) | (-0.11) | (-0.02) | (0.56) | (0.55) |
| 9 | 0.02*** | 0.76*** | -0.06 | -0.08 | -0.12 | -0.13 | -0.13 | 0.02*** | 0.85*** | 0.05 | 0.09 | 0.06 | -0.01 | -0.01 |
| | (19.52) | (3.15) | (-0.66) | (-1.00) | (-1.33) | (-1.14) | (-1.15) | (19.29) | (3.77) | (0.76) | (1.28) | (0.81) | (-0.12) | (-0.12) |
| H | 0.04*** | 0.63*** | -0.09 | -0.22** | -0.25*** | -0.30*** | -0.30*** | 0.04*** | 0.70*** | -0.05 | -0.16** | -0.20*** | -0.25*** | -0.25*** |
| | (18.75) | (2.99) | (-1.04) | (-2.61) | (-3.19) | (-2.99) | (-2.99) | (17.80) | (3.37) | (-0.69) | (-2.16) | (-2.80) | (-2.92) | (-2.98) |
| H-L | 0.05*** | -0.72*** | -0.62*** | -0.69*** | -0.82*** | -0.81*** | -0.81*** | 0.05*** | -0.53*** | -0.46*** | -0.40** | -0.54*** | -0.63*** | -0.63*** |
| | (17.49) | (-3.91) | (-3.65) | (-3.42) | (-4.31) | (-2.98) | (-3.39) | (16.42) | (-3.01) | (-3.11) | (-2.20) | (-3.22) | (-2.71) | (-3.26) |

Table A.5: **Performance of decile portfolios sorted on the net gamma exposure:** This table reports the performance of decile portfolios formed on the basis of the net gamma exposure (Γ), which measures the total outstanding gamma divided by the market capitalization. At the end of month $t$ we sort stocks into ten portfolios based on their Γ, and hold this portfolio during month $t+1$. Panel A (B) presents the results for value-weighted portfolios whereby the breakpoints are based on the full sample (NYSE universe). Stocks with prices above $5 and microcaps as of the portfolio formation are excluded. We report the average Γ, the return ("R") in percentages, the Fama-French-Carhart four-factor alpha ("$\alpha_{3FM}$"), the Fama-French-Carhart five-factor alpha ("$\alpha_{5F}$"), the Fama-French-Carhart six-factor alpha ("$\alpha_{5FM}$"), Hou, Xue, and Zhang's extended q-factor model alpha ("$\alpha_{5Q}$"), and augmented with momentum ("$\alpha_{5QM}$") for each portfolio. The row labeled "L-H" is the self-financing high-minus-low portfolio, which reports the difference in between portfolio H and portfolio L. The sample consists of stocks listed on NYSE/AMEX/NASDAQ for the period between January 1996 and December 2021 with share code 10 or 11. Newey-West t-statistics are reported between parentheses. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level. The sample runs from February 1996 until December 2021.

| | **Panel A: Full sample breakpoints** | | | | | | | **Panel B: NYSE-breakpoints** | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Γ | $R$ | $\alpha_{3FM}$ | $\alpha_{5F}$ | $\alpha_{5FM}$ | $\alpha_{Q5}$ | $\alpha_{Q5M}$ | Γ | $R$ | $\alpha_{3FM}$ | $\alpha_{5F}$ | $\alpha_{5FM}$ | $\alpha_{Q5}$ | $\alpha_{Q5M}$ |
| L | -0.10*** | 1.45*** | 0.64*** | 0.53*** | 0.65*** | 0.59*** | 0.59*** | -0.10*** | 1.44*** | 0.63*** | 0.53*** | 0.65*** | 0.58*** | 0.58*** |
| | (-8.63) | (5.11) | (4.18) | (3.80) | (4.63) | (3.27) | (3.89) | (-8.63) | (5.00) | (4.08) | (3.64) | (4.57) | (3.15) | (3.78) |
| 2 | -0.01*** | 1.16*** | 0.34*** | 0.16 | 0.22** | 0.23** | 0.23** | -0.01*** | 1.22*** | 0.41*** | 0.20* | 0.26** | 0.27** | 0.27*** |
| | (-6.28) | (4.83) | (3.07) | (1.23) | (2.08) | (1.97) | (2.51) | (-6.25) | (5.31) | (3.27) | (1.72) | (2.59) | (2.52) | (3.30) |
| 3 | 0.00*** | 1.15*** | 0.38*** | 0.25 | 0.32** | 0.32* | 0.32** | 0.00*** | 1.12*** | 0.33** | 0.15 | 0.23* | 0.19 | 0.20 |
| | (2.95) | (4.74) | (2.67) | (1.52) | (2.29) | (1.86) | (2.19) | (2.68) | (4.65) | (2.48) | (0.97) | (1.79) | (1.17) | (1.44) |
| 4 | 0.01*** | 0.91*** | 0.11 | -0.03 | -0.01 | -0.03 | -0.03 | 0.01*** | 0.96*** | 0.15* | 0.06 | 0.08 | 0.05 | 0.05 |
| | (13.06) | (4.03) | (1.07) | (-0.29) | (-0.08) | (-0.32) | (-0.34) | (13.46) | (4.17) | (1.80) | (0.77) | (1.07) | (0.53) | (0.55) |
| 5 | 0.02*** | 0.88*** | 0.09 | 0.04 | 0.07 | 0.09 | 0.09 | 0.02*** | 1.01*** | 0.22** | 0.11 | 0.14 | 0.17 | 0.17 |
| | (18.17) | (3.51) | (0.87) | (0.37) | (0.72) | (0.87) | (0.88) | (17.40) | (4.43) | (2.05) | (0.83) | (1.12) | (1.26) | (1.33) |
| 6 | 0.04*** | 1.03*** | 0.22* | 0.16 | 0.18 | 0.24 | 0.24* | 0.04*** | 0.94*** | 0.17* | 0.09 | 0.13 | 0.22* | 0.22* |
| | (20.06) | (4.27) | (1.94) | (1.45) | (1.47) | (1.63) | (1.77) | (17.86) | (3.78) | (1.68) | (0.87) | (1.30) | (1.71) | (1.94) |
| 7 | 0.06*** | 0.84*** | 0.05 | 0.06 | 0.07 | 0.02 | 0.02 | 0.06*** | 0.78*** | -0.01 | -0.03 | -0.02 | -0.02 | -0.02 |
| | (19.34) | (3.42) | (0.54) | (0.67) | (0.72) | (0.24) | (0.24) | (16.81) | (3.21) | (-0.16) | (-0.31) | (-0.27) | (-0.20) | (-0.20) |
| 8 | 0.10*** | 0.70** | -0.13 | -0.14 | -0.15* | -0.15 | -0.15 | 0.10*** | 0.87*** | 0.06 | 0.02 | 0.02 | 0.00 | 0.00 |
| | (17.73) | (2.60) | (-1.59) | (-1.50) | (-1.68) | (-1.48) | (-1.53) | (15.39) | (3.49) | (0.69) | (0.24) | (0.23) | (0.02) | (0.02) |
| 9 | 0.17*** | 0.74** | -0.14 | -0.00 | -0.07 | -0.12 | -0.12 | 0.16*** | 0.73*** | -0.11 | -0.05 | -0.09 | -0.14 | -0.14 |
| | (15.88) | (2.74) | (-1.60) | (-0.03) | (-0.84) | (-1.11) | (-1.39) | (13.90) | (2.71) | (-1.13) | (-0.40) | (-0.86) | (-1.19) | (-1.31) |
| H | 0.40*** | 0.78** | -0.23 | 0.01 | -0.08 | -0.17 | -0.17 | 0.38*** | 0.69** | -0.28** | -0.08 | -0.16 | -0.21 | -0.21* |
| | (13.24) | (2.17) | (-1.55) | (0.05) | (-0.47) | (-0.97) | (-1.17) | (12.25) | (2.19) | (-2.54) | (-0.63) | (-1.46) | (-1.53) | (-1.87) |
| H-L | 0.49*** | -0.67** | -0.88*** | -0.52** | -0.73*** | -0.76** | -0.76*** | 0.48*** | -0.75*** | -0.91*** | -0.62*** | -0.82*** | -0.79*** | -0.79*** |
| | (8.64) | (-2.45) | (-3.80) | (-2.00) | (-3.31) | (-2.53) | (-3.22) | (7.92) | (-3.02) | (-4.13) | (-2.72) | (-4.02) | (-2.78) | (-3.53) |

# Chapter 5

# Summary

This dissertation consists of four empirical essays in asset pricing and financial markets, ranging from macro-economic nowcasting to systematic high-frequency return patterns. Each essay establishes new empirical patterns, which provides deeper understanding on the behaviour of asset prices.

Chapter 2 studies how asset prices, at the index-level, respond to macroeconomic news. Practically every day, macroeconomic news is released, containing information on the state of the economy. The unexpected component of this news causes investors to update their expectations, and are key inputs in investment decision-making processes. However, only a few studies comprehensively study the effect of macroeconomic surprises and asset pricing.

We contribute to this set of literature by focusing on three key dimensions. First, we utilize a large panel (over 200+) of macroeconomic surprises in four major regions (U.S., U.K., Japan, and Europe). This allows to model for the daily macroeconomic news flow that is well-followed by investors. These macroeconomic surprises are used to develop a novel aggregate surprise index, in real-time, and on the daily frequency. Second, we study the behaviour of macroeconomic surprises. Lastly, we consider how macroeconomic surprises affect asset prices across major asset class.

From an econometric perspective, our modelling technique is able to tackle several challenges that often occur in macro-economic nowcasting. Our methodology is able to aggregate a large number of releases into a single metric (i), focus on well-interpretable categories of economic information (ii), rely on real-time information (iii), and offer a high-frequency measure of macro-economic surprises (iv). Moreover, the method is simple to implement, and is able to handle data with different release dates, and frequencies. From a behavioural perspective, we establish that economic surprises do not follow a random walk, but rather exhibit sizable short-term autocorrelation. We term this empirical pattern *"economic surprise momentum"*, and caused by (i) underreaction in the consensus forecast within the surprise series, and novel to the literature, (ii) underreaction in forecasts of other series. Our results suggest that economic forecasters and investors show behaviour inconsistent with economic rationality. Lastly, we examine the predictability of returns across all major asset classes and across major markets. we find that asset returns are strongly predicted by macroeconomic surprises. Simple investment strategies yield sizable Sharpe ratios and CAPM alphas.

In chapter 3, we apply the concept of *"Non-Standard Errors"* to the factor investing literature. Non-standard errors capture the variation in outcomes induced by methodological decisions across researchers, in addition to statistical uncertainty. we specifically focus on portfolio sorting, which is the most used methodology within empirical asset pricing. As researchers face a number of design choices when engaging in portfolio sorting, the exact procedure is not uniform across studies, which might lead to considerable variation in outcomes.

Our first contribution is that we conduct a large-scale literature review (over 300+ papers) to take stock of the variation in methodological choices made in portfolio sorts. After our mapping, we document that 11 options are highly important, leaving much room for the researcher to potentially cherry-pick the best choices. Furthermore, we find that the choices among these options are quite dispersed among researchers. For example, using value-weighting instead of equal-weighting returns in portfolios occurs approximately 50-50. we also find that, historically, much of the choices that are being made tend to be choices that inflate performance metrics such as alphas and Sharpe ratios.

Second, we construct well-known factors and factor models using all possible combinations of the 11 documented choices, leaving 2048 ($2^{11}$) versions of each factor. We find that factors exhibit large variation in Sharpe ratios within our set of possible construction methods. To illustrate, the value factor yields a Sharpe ratio between 0.15 and 1.24, depending on the choices being made. Furthermore, we find that the ratio of non-standard errors to standard errors is typically above 1. This implies that the former is equally important as the latter in terms of economic magnitude.

Third, we perform model comparison tests among different factor models using the 2048 different combinations. We find that there is no single "winning" model, once the whole range of possibilities is considered. For example, the Barillas-Shanken 6-factor model and the Daniel-Hirshleifer-Sun 3-factor model have a 50-50 probability of being the best performing factor model.

Overall, this chapter concludes that factor design choices are essential, and recommends the use of multiple construction methods to reduce potential p-hacking and data-mining. For future studies, one way forward is to consider these choices in a "specification check", in which the distribution of results from multiple methodological possibilities is reported. Another suggestion is to be more uniform in portfolio sorting choices, using conservative and prudent research designs.

Chapter 4 introduces the net gamma exposure as an important predictor for the cross-section of equity returns in the U.S. Option trading has been increasing volume-wise in the past two decades. This essay raises a key question: does option trading affect the price dynamics of the underlying asset? We argue that, by mandate, option market makers hedge themselves to remain delta-neutral. Such hedging behaviour can have large impact on asset prices, depending on the size of such hedging flow. The amount that option market makers need to hedge to remain delta-neutral depends on gamma, which measures how much delta moves when the underlying price changes. we directly proxy for this amount by constructing a measure of net gamma exposure (NGE) at the stock-level, which equals the gamma-weighted sum of open interest across the option chain on that stock, scaled by the average dollar trading volume in the past 21 business days. This measure can be interpreted as the amount that needs to be hedged, given a 1% in the price of the underlying.

Empirically, an univariate sort on NGE results in a decile return spread of 10% (annualized), indicating that stocks with lower NGE tend to outperform stocks with high NGE. In a regression-based approach, we also find that NGE is a strong negative predictor of next month's stock return. Using a wide-range of specification checks (as advocated in Chapter 2), this negative predictability remains robustly negative and statistically significant.

Why is NGE negatively priced in the cross-section of equity returns? We argue that risk-averse investors demand additional compensation in the form of higher returns to hold stocks with negative gamma exposure. When NGE is negative (positive), delta decreases (increases) when the underlying price increases. As such, market makers are required to buy (sell) the underlying more aggressively after an increase in the underlying price, resulting in additional positive (negative) price pressure. Thus, the initial price movement is amplified when NGE is negative, and the corresponding stock volatility is higher. We empirically confirm that NGE negatively and significantly predict future realized volatility.

In the last essay, in chapter **??**, we study intraday returns in the cross-section of U.S. equity returns. Specifically, we document a return pattern that we name "Intraday Reversal": stocks with high returns during the trading day until 3:30 pm (ROD) tend to underperform in the last half hour (LH) compared to stocks with low returns during the trading session (except the last half hour).

We show that this predictability is mainly concentrated among negative ROD observations. A range of robustness tests show that the results still hold in, amongst others, a sample consisting of liquid and large firms, and consistently over time. The relationship between ROD and LH is not driven by other well-known intraday patterns, such as the tug-of-war effect and the intraday seasonality effect.

Previous literature argues that gamma-related order flow affects end-of-day price dynamics, and might cause intraday reversal or momentum, depending on the sign of gamma. However, we show that intraday reversal is present for non-optionable stocks, and stocks with little price pressure from gamma-related flows. This implies that intraday reversal is not only driven by hedge rebalancing of option market makers.

We argue that intraday reversal is rather driven by retail investors that buy stocks that decreased in value intra-day ('dip-buying behaviour'), and short-sellers that close positions when there is a intraday price decrease ('profit-taking'). We show that ROD negatively predicts retail order imbalance in the last half hour, indicating increased buy pressure from retail investors, consistent with our hypothesis. Additionally, the retail order imbalance reverses the next morning, further supporting the notion of temporary price pressure. In addition, we show that intraday reversal is weaker for stocks with high institutional ownership (low retail ownership and trading). Likewise, we document that ROD negatively predicts short volume in the last half hour, indicating more short selling activity when ROD is low. In addition, I compare intraday reversal with a seemingly contrary effect: market intraday momentum. At the market-level it is shown that ROD positively predicts LH returns, which is opposite to the intraday reversal that we document at the stock-level. After controlling for market-level momentum, our reversal effect still persists.

The evolution of asset pricing theory is a fascinating journey spanning many decades, and molded by a variety of economic and financial insights. Traditionally, in the 60s and 70s, asset pricing theory was formed under the assumption of full information and rational expectations. These assumptions gave rise to the existence of a "homo economicus" in financial theory. This "homo economicus" takes unbiased and impartial actions in order to maximize his own utility. Although appealing and elegant, these assumptions do not accurately capture the behaviour of individuals, nor markets. Behavioural finance emerged, in the 80s and 90s, to highlight the psychological biases and irrational behaviour of human beings, leading to deviations from rationality and market inefficiencies. The 90s until present day have been characterized by researchers discovering various empirical anomalies that seems to challenge the assumptions of traditional asset pricing theories. The current state of asset pricing aims to refine theory to match the various stylized facts documented by empiricists, as well as incorporating behavioural aspects. Ultimately, modern asset pricing aims to provide better and more realistic explanations for the complexities of financial markets. The field continues to evolve, with ongoing debates, novel stylized facts, and new financial challenges.

This dissertation examines some of these recent challenges in asset pricing research, thereby documenting novel stylized facts, and contributing to several important ongoing debates. Overall, what can academic researchers and practitioners learn from the findings and implications of this dissertation? All results presented here, one way or another, suggests that individuals and markets do not behave according to the traditional finance paradigm. For example, we have shown in chapter 2 that professional macroeconomic forecasters are systematically biased, and under-react to new information, which is inconsistent with the full information and rational expectations assumptions. Chapter 3 offers guidance on prudent research in empirical asset pricing research in order to successfully identify asset pricing anomalies. Chapter 4 documents a novel asset pricing anomaly that shows that information contained in option trades is not correctly priced in by investors, thereby challenging market efficiency theories. Chapter 5 documents an intraday patterns whereby retail traders are contrarian traders in the last half-hour of the trading session, thereby challenging investor homogeneity which is assumed in traditional asset pricing models.

The recognition that financial rationality assumptions do not hold has significant implications for academic researchers, practitioners, and policymakers. For academics, the failure of the traditional finance paradigm asks for continuous development of more realistic models that effectively capture human behaviour and financial market dynamics. To the practitioner: acknowledging that investors may not act rationally helps to design more effective risk management strategies. Behavioral factors can contribute to market volatility, which should be considered in risk management frameworks. In addition, asset pricing anomalies can be exploited by practitioners via various types of investment strategies. Lastly, financial advisors can integrate behavioural insights in order to provide more tailored and effective financial advice. Policymakers can consider behavioral insights when designing regulations to dampen market distortions caused by irrational behavior. This might involve implementing measures that encourage more prudent decision-making. Lastly, the failure of the traditional financial paradigm also holds valuable insights to the layperson. Studying simple principles in behavioural finance as part of financial education (next to traditional economics) equips the layperson with a broader toolkit when making financial decisions.

# Samenvatting in het Nederlands

Dit proefschrift bestaat uit vier empirische essays over financiele markten en het prijzen van financiele producten, variërend van macroeconomische "nowcasting" tot systematische rendementspatronen op hoge frequentie. Elk essay documenteert nieuwe empirische patronen die een dieper inzicht verschaffen in het gedrag van prijzen van financiele producten.

Hoofdstuk 2 bestudeert hoe indexprijzen reageren op macro-economische verrassingen. Vrijwel dagelijks komt er macro-economisch nieuws naar buiten met daarin informatie over de stand van de economie. De onverwachte component van dit nieuws zorgt ervoor dat beleggers hun verwachtingen bijstellen en vormt een belangrijke input in besluitvormingsprocessen voor investeringen. Toch bestuderen slechts enkele studies het effect van macro-economische verrassingen op indexprijzen.

Ik draag bij aan deze reeks literatuur door me te concentreren op drie belangrijke dimensies. Ten eerste maak ik gebruik van een dataset (meer dan 200 reeksen) van macro-economische verrassingen in vier grote regio's (VS, VK, Japan en Europa). Dit maakt het mogelijk om de dagelijkse stroom van nieuws te modelleren. Deze macro-economische verrassingen worden gebruikt om een nieuwe geaggregeerde verrassingsindex te ontwikkelen, in realtime en op een dagelijkse frequentie. Ten tweede bestudeer ik het gedrag van macro-economische verrassingen. Ten slotte ga ik na hoe macro-economische verrassingen de indexprijzen beïnvloeden in de belangrijkste klassen van financiële producten.

Vanuit een econometrisch perspectief is onze modelleringstechniek in staat om verschillende uitdagingen aan te gaan die vaak voorkomen bij macro-economische nowcasting. Onze methodologie kan een groot aantal nieuwsuitgiftes samenvoegen tot een enkele metriek (i), zich concentreren op goed interpreteerbare categorieën van economische informatie (ii), realtime informatie gebruiken (iii), en een hoogfrequente meting bieden van macro-economische verrassingen (iv). Bovendien is de methode eenvoudig te implementeren en kan deze omgaan met data die verschillende frequenties hebben en op verschillende momenten door de tijd vrijkomen.

Bovendien stel ik vanuit een gedragsperspectief vast dat economische verrassingen niet het gevolg zijn van een willekeurige wandeling, maar eerder autocorrelatie vertonen op korte termijn. We noemen dit empirische patroon het "economisch verrassingsmomentum", en het wordt veroorzaakt door (i) onderreactie in de consensusprognose binnen de verrassingsreeks en, nieuw in de literatuur, (ii) onderreactie in prognoses tussen reeksen.

Tot slot onderzoek ik de voorspelbaarheid van indexrendementen voor de belangrijkste markten en soorten financiële producten. Ik ontdek dat activarendementen sterk worden voorspeld door macro-economische verrassingen. Eenvoudige beleggingsstrategieën leveren aanzienlijke Sharpe-ratio's en CAPM-alpha's op.

In hoofdstuk 3 pas ik het concept van "Non-Standard Errors" toe op de literatuur over factorbeleggen. Non-Standard Errors geven de variatie in uitkomsten weer, bovenop de statistische onzekerheid, veroorzaakt door variatie in methodologische beslissingen tussen onderzoekers. Ik richt me specifiek op het sorteren van portefeuilles, de meest gebruikte methodologie binnen empirische asset pricing. Aangezien onderzoekers bij het sorteren van portefeuilles met een flink aantal keuzes worden geconfronteerd, is de exacte procedure niet uniform voor alle onderzoeken, wat kan leiden tot aanzienlijke variatie in uitkomsten.

Mijn eerste bijdrage is dat ik een grootschalig literatuuronderzoek uitvoer (met meer dan 300 artikelen) om de variatie in methodologische keuzes die bij portefeuille sortering worden gemaakt te inventariseren. Ik stel vast dat 11 keuzes zeer belangrijk zijn, waardoor er veel ruimte overblijft voor de onderzoeker om mogelijk de beste opties te kiezen. Bovendien merk ik op dat de keuzes tussen deze opties verdeeld zijn onder onderzoekers. Bijvoorbeeld, het gebruik van 1/N-weging of een markt-weging in portefeuilles komt ongeveer 50-50 voor. Ik merk ook op dat, historisch gezien, veel van de gemaakte keuzes vaak de prestatiestatistieken zoals alpha's en Sharpe-ratio's opblazen.

Ten tweede construeer ik bekende factoren en factormodellen met alle mogelijke combinaties van de 11 gedocumenteerde keuzes, waardoor er 2048 ($2^{11}$) versies worden gemaakt van iedere factor. We vinden dat factoren een grote variatie vertonen in Sharpe-ratio's binnen onze reeks mogelijke constructiemethoden. Ter illustratie: de value-factor levert een Sharpe-ratio op tussen 0,15 en 1,24, afhankelijk van de gemaakte keuzes. Verder vinden we dat de verhouding van niet-standaardfouten ten opzichte van standaardfouten doorgaans boven de 1 ligt. Dit impliceert dat niet-standaardfouten even belangrijk zijn als de laatste in termen van economische omvang.

Ten derde vergelijk ik de prestaties van de verschillende modellen met behulp van de 2048 verschillende combinaties. Onze resultaten laten zien dat er niet één "winnend" model is, als het hele scala aan mogelijkheden eenmaal is overwogen. Het Barillas-Shanken 6-factorenmodel en het Daniel-Hirshleifer-Sun 3-factorenmodel hebben bijvoorbeeld een kans van 50-50 om het best presterende factormodel te zijn.

Dit hoofdstuk concludeert dat keuzes voor het ontwerpen van factoren essentieel zijn en beveelt het gebruik van meerdere constructiemethoden aan om potentiële p-hacking en data-mining te verminderen. Voor toekomstig onderzoek raden wij aan om een uitgebreide specificatiecheck uit te voeren waarbij de verdeling van alle mogelijke resultaten wordt getoond op basis van de verschillende mogelijke keuzes. Een andere suggestie is om meer uniform te zijn in de keuzes voor het sorteren van portefeuilles, met behulp van conservatieve onderzoeksontwerpen.

Hoofdstuk 4 introduceert de netto gamma blootstelling (NGE) als een belangrijke voorspeller van aandelenrendementen in de VS. De handel in opties is de afgelopen twee decennia sterk toegenomen in volume. Dit hoofdstuk stelt een belangrijke vraag: heeft optiehandel invloed op de prijsdynamiek van het onderliggende product? Ik redeneer dat optiehandelaren risico's indekken om delta-neutraal te blijven, omdat dat hun mandaat is. Een dergelijke indekking kan een grote impact hebben op de prijzen van aandelen, afhankelijk van de omvang van de indekking. Hoeveel een optiehandelaar moet afdekken om delta-neutraal te blijven, hangt af van de gamma. De gamma van een optie meet hoeveel de delta verandert wanneer de onderliggende prijs verandert. In dit hoofdstuk construeer ik een maatstaf die meet hoeveel er door optiehandelaren moet worden afgedekt voor ieder aandeel op een dag. Mijn maatstaf is de gamma-gewogen som van de openstaande hoeveelheid opties voor een gegeven aandeel, geschaald door het gemiddelde handelsvolume in dollars in de afgelopen 21 werkdagen. Deze maatstaf kan worden geïnterpreteerd als het bedrag dat moet worden afgedekt, gegeven een prijsverandering van 1% in de onderliggende waarde.

Ik sorteer alle aandelen op basis van deze maatstaf in tien groepen. Gemiddeld genomen verdienen aandelen in de laagste groep 10% meer (op jaarbasis) dan aandelen in de hoogste groep. Aan de hand van een regressieraamwerk, vind ik ook dat mijn maatstaf een sterke negatieve voorspeller is van het aandelenrendement van volgende maand. Met behulp van een breed scala aan specificatiecontroles (zoals aanbevolen in hoofdstuk 2), blijft deze negatieve voorspelbaarheid robuust negatief en statistisch significant.

De gevonden negatieve relatie kan worden verklaard door beleggers die risicomijdend zijn en een compensatie eisen in de vorm van hogere rendementen om aandelen met een negatieve gamma-blootstelling aan te houden. Wanneer de blootstelling negatief (positief) is, neemt de delta af (toe) wanneer de onderliggende prijs stijgt. Optiehandelaren moeten dan aandelen agressiever bijkopen (verkopen) na een stijging van de onderliggende prijs, wat leidt tot additionele positieve (negatieve) prijsdruk. Dit versterkt de initiële prijsbeweging wanneer de blootstelling negatief is, en daarmee ook de volatiliteit van een aandeel. Ik lever bewijs voor deze hypothese door te laten zien dat de gamma-blootstelling ook toekomstige gerealiseerde volatiliteit negatief voorspelt. Deze bevinding ondersteunt het idee dat de negatieve relatie tussen gamma-blootstelling en rendementen te wijten kan zijn aan prijsdruk en verhoogde volatiliteit in aandelen met negatieve blootstelling.

In mijn laatste onderzoek, in hoofdstuk **??**, bestudeer ik intradag-rendementen voor Amerikaanse aandelen. Ik stel een nieuwe patroon vast dat ik "intraday reversal" noem: aandelen met een hoog rendement gemeten tussen gister bij sluit en vandaag tot en met 15:30 (dit noem ik het ROD interval), hebben de neiging om slechter te presteren in het laatste half uur van een handelsdag in vergelijking met aandelen die het juist slecht deden.

Ik laat zien dat deze voorspelbaarheid voornamelijk geconcentreerd is bij negatieve ROD-waarnemingen. Uit een reeks robuustheidstesten blijkt dat de resultaten nog steeds standhouden in onder meer een steekproef bestaande uit liquide en grote bedrijven, en dat ze consistent zijn over tijd. De relatie tussen ROD en LH wordt niet gedreven door andere bekende intradag-patronen, zoals het touwtrek-effect ("tug-of-war") en intradag-seizoenseffecten.

Eerdere literatuur betoogt dat gamma-gerelateerde handelsstromen de prijsdynamiek aan het einde van de dag beïnvloeden en intraday-reversal of momentum kunnen veroorzaken, afhankelijk van het teken van gamma. Ik laat echter zien dat intraday reversal aanwezig is voor aandelen zonder beschikbare opties en aandelen met weinig prijsdruk door gamma-gerelateerde stromen. Dit impliceert dat intraday reversal niet alleen wordt aangestuurd door herbalancering van optiemarktmakers. Dit wijst erop dat andere factoren dan alleen gamma-gerelateerde stromen een rol spelen bij het veroorzaken van intraday reversal en dat het fenomeen breder is dan alleen gerelateerd aan optiemarktactiviteiten.

Ik beargumenteer dat intraday reversal eerder wordt gedreven door particuliere beleggers die aandelen kopen die intraday in waarde zijn gedaald ('dip-buying-gedrag'), en short-sellers die posities sluiten wanneer ('winstneming'). Ik laat zien dat ROD een negatieve voorspeller is van de onbalans in retail orders in het afgelopen half uur, wat wijst op een verhoogde koopdruk van retailbeleggers, in overeenstemming met mijn hypothese. Bovendien keert de onbalans in de retailorders de volgende ochtend om, in overeenstemming met tijdelijke prijsdruk. Daarnaast laat ik zien dat intraday reversal zwakker is voor aandelen die vooral in bezit zijn van institutionele investeerders. Evenzo documenteer ik dat ROD de short-selling volume in het laatste half uur negatief voorspelt, wat wijst op meer short selling-activiteit wanneer ROD laag is. Bovendien vergelijk ik intraday reversal met een schijnbaar tegengesteld effect: intraday momentum op het index-niveau. Op marktniveau wordt aangetoond dat ROD de LH-rendementen positief voorspelt, wat tegengesteld is aan de intraday reversal die ik documenteer op aandelenniveau. Na correctie voor momentum op marktniveau houdt onze reversal effect nog steeds aan.

De evolutie van financiele theorie is een reis die vele decennia omspant en wordt gevormd door verschillende economische en financiële inzichten. Traditioneel, in de jaren 60 en 70, werd financiele theorie ontwikkeld onder de veronderstelling van volledige informatie en rationele verwachtingen. Deze aannames hebben geleid tot het bestaan van de "homo economicus" in financiële theorie. Deze "homo economicus" handelt onbevooroordeeld en onpartijdig om zijn eigen nut te maximaliseren. Hoewel aantrekkelijk en elegant, deze aannames beschrijven het gedrag van individuen en financiele markten niet nauwkeurig genoeg. 'Behavioural finance" ontstondt in de jaren 80 en 90 om de cognitieve fouten en irrationeel gedrag van mensen in kaart te brengen, welke afwijkt van de rationaliteitsaannames uit de jaren 60. De jaren 90 tot heden worden gekenmerkt door onderzoekers die verschillende empirische anomalieën hebben ontdekt die de veronderstellingen van traditionele theorieën uit dagen. Tegenwoordig streven onderzoekers ernaar de financiele theorie te verfijnen om overeen te komen met de verschillende feiten die door empirici zijn gedocumenteerd, evenals het opnemen van gedragsaspecten in de theorie. Uiteindelijk streeft moderne financiele theorie naar betere en realistischere verklaringen voor de complexiteit van financiële markten. Het vakgebied blijft evolueren, met doorlopende debatten, nieuwe empirische feiten en nieuwe financiële uitdagingen.

Deze dissertatie onderzoekt enkele van deze recente uitdagingen en documenteert daarbij nieuwe empirische feiten en draagt bij aan verschillende huidige discussies. Wat kunnen academische onderzoekers en professionals leren van de bevindingen en implicaties van deze dissertatie? Alle gepresenteerde resultaten suggereren, linksom of rechtsom, dat individuen en markten zich niet gedragen volgens het traditionele financiële paradigma. Zo hebben we in hoofdstuk 2 aangetoond dat professionele macro-economische voorspellers systematisch bevooroordeeld zijn en onderreageren op nieuwe informatie, wat niet overeenkomt met de veronderstellingen van volledige informatie en rationele verwachtingen. Hoofdstuk 3 biedt richtlijnen voor zorgvuldig onderzoek om succesvol anomalieën in financiele producten te kunnen identificeren. Hoofdstuk 4 documenteert een nieuw anomalie in aandelen dat aantoont dat informatie uit optietransacties, niet correct wordt geprijsd door investeerders, waarmee de theorieën van marktefficiëntie worden uitgedaagd. Hoofdstuk 5 documenteert patronen gedurende de dag waarbij particuliere handelaren contrair handelen in het laatste halfuur van de handelssessie, waarmee de homogeniteit van beleggers wordt uitgedaagd die wordt verondersteld in traditionele financiele theorie.

Het besef dat financiële rationaliteitsaannames niet van toepassing zijn, heeft significante implicaties voor academische onderzoekers, professionele beleggers en beleidsmakers. Voor academici vraagt het falen van het traditionele financiële paradigma om voortdurende ontwikkeling van realistischere modellen die effectief menselijk gedrag en de dynamiek op financiële markten vastleggen. Voor professionele beleggers: het erkennen dat investeerders mogelijk niet rationeel handelen, helpt bij het ontwerpen van effectievere risicobeheerstrategieën. Gedragsfactoren kunnen bijdragen aan marktvolatiliteit, wat moet worden meegenomen in het beheren van risico's. Daarnaast kunnen beleggingsprofessionals gebruikmaken van anomalieën in financiele instrumenten via verschillende soorten investeringsstrategieën. Bovendien kunnen financiële adviseurs gedragsinzichten integreren om meer op maat gemaakt en effectief financieel advies te bieden aan clienten. Beleidsmakers kunnen gedragsinzichten overwegen bij het ontwerpen van reguleringen om marktverstoringen als gevolg van irrationeel gedrag te verminderen. Dit kan het implementeren van maatregelen omvatten die meer prudente besluitvorming bevorderen. Ten slotte bevatten de beperkingen van het traditionele financiële paradigma ook waardevolle inzichten voor individuen zonder financiele expertise. Het bestuderen van eenvoudige principes in gedragsfinanciën als onderdeel van financiële educatie (naast traditionele economie) voorziet deze individuen van een breder blik bij het nemen van financiële beslissingen.

# About the Author

Amar Soebhag was born on 14 February 1996 in Rotterdam, the Netherlands. He holds a bachelor degree in Economics and Business Economics (summa cum laude) from the Erasmus University Rotterdam, where he graduated in 2016. Furthermore, he holds a master degree in Economics with a specialization in Financial Economics (cum laude, 2017) from the Erasmus University Rotterdam. To prepare for his PhD, Amar enrolled into the Research master programme (MPhil) of the Tinbergen Institute, with a specialization in finance and econometrics. He completed the Mphil programme in 2019, and afterwards pursued his PhD at the Erasmus School of Economics under the supervision of Prof. Dr. Guido Baltussen, and Prof. Dr. Patrick Verwijmeren.

Amar specializes in systematic investment strategies, macroeconomic forecasting, high-frequency strategies, and machine learning. He presented his work at leading international conferences, such as the annual meeting of the European Financial Association (2022), and the European meeting of the Financial Management Association (2023). In addition, Amar has (co-)supervised over 100 bachelor and master theses over the period 2019-2023. He taught seminars in Behavioural Investing at the master-level at the Erasmus School of Economics. Amar is currently employed at Robeco Quantitative Investments as a Quant Researcher in the Quantitative Fixed Income Team, where he applies his expert knowledge on quantitative investing.

# Portfolio/CV

### Published papers

1. **Reducing socioeconomic health inequalities? A questionnaire study of majorization and invariance conditions:** Joint with Kirsten Rohde and Tom van Ourti. *Journal of Health Economics, volume 90, July 2023.*

### Working papers

1. **Caught by Surprise: How Markets Respond to Macroeconomic News**. Joint with Guido Baltussen.
2. **Non-Standard Errors in Asset Pricing: Mind Your Sorts**. Joint with Bart van Vliet and Patrick Verwijmeren
3. **Option Gamma and Stock Returns**.
4. **Intraday Reversal**. Joint with Guido Baltussen.

### Work in progress

1. **The low volatility factor:** Joint with Guido Baltussen and Pim van Vliet
2. **Mispricing, short-sellers and intraday returns:** Joint with Esad Smaljbegovic

### Teaching

1. Seminar Behavioural Investing: this seminar is taught at the Erasmus School of Economics within the MSc. programme "Financial Economics". The following tasks were involved: teaching and supervising +/- 25 students per year. Organizing, reviewing and grading group projects on empirical asset pricing.
2. Supervising bachelor and master theses: (co-)supervising over 100 students between 2019-2023 in writing their theses in order to graduate, leading defense ceremonies, and grading the theses.

3. Finance 1: this is a compulsory course in all bachelor programmes offered by the Erasmus School of Economics, and consists of over 900+ students per year. The following tasks were involved: hiring/recruiting and guiding teaching assistants, and planning the schedule for the course.

**PhD courses**

I have completed over 100 ECTs in course work during the Research master programme at the Tinbergen Institute. I have followed the following courses:

| Course Name | Year | Block | Field | ECTS |
|---|---|---|---|---|
| Principles of Programming | 1st year | 1 | Statistics/Econometrics | 1 |
| Microeconomics I (Individual Decision-making) | 1st year | 1 | Microeconomics | 4 |
| Asymptotic Theory | 1st year | 1 | Statistics/Econometrics | 4 |
| Mathematics | 1st year | 1 | Statistics/Econometrics | 4 |
| Advanced Econometrics I | 1st year | 2 | Statistics/Econometrics | 4 |
| Macroeconomics I (Neoclassical models) | 1st year | 2 | Macroeconomics | 4 |
| Microeconomics II (Game theory) | 1st year | 2 | Microeconomics | 4 |
| Advanced Econometrics II | 1st year | 3 | Statistics/Econometrics | 4 |
| Macroeconomics II (Macroeconomic Policy) | 1st year | 3 | Macroeconomics | 4 |
| Asset pricing | 1st year | 3 | Finance | 4 |
| Advanced Econometrics III | 1st year | 4 | Statistics/Econometrics | 4 |
| Microeconomics IV (Behavioural Economics) | 1st year | 4 | Microeconomics | 4 |
| Corporate Finance | 1st year | 4 | Finance | 4 |
| Macroeconomics IV (Financial Frictions) | 1st year | 5 | Macroeconomics | 4 |
| Experimental Economics | 1st year | 5 | Microeconomics | 3 |
| International Economics | 1st year | 5 | Macroeconomics | 3 |
| Behavioural Finance | 1st year | 5 | Finance | 3 |
| Econometrics Lectures (Serena Ng) | 1st year | 5 | Statistics/Econometrics | 3 |
| Academic writing | 1st year | 5 | - | 1 |
| Banking | 2nd year | 1 | Finance | 3 |
| Applied Microeconometrics | 2nd year | 1 | Statistics/Econometrics | 3 |
| Continuous Time Asset Pricing | 2nd year | 2 | Finance | 3 |
| Bayesian Econometrics | 2nd year | 2 | Statistics/Econometrics | 3 |
| Empirical Asset Pricing | 2nd year | 5 | Finance | 3 |
| Econometrics Lecture Series (Guido Imbens) | 2nd year | 5 | Statistics/Econometrics | 3 |
| Master thesis | 2nd year | 3-5 | Finance | 30 |

**Tinbergen Dissertation Series**

Tinbergen Institute (TI) is the graduate school and research institute operated jointly by the Schools of Economics and Econometrics of Erasmus University Rotterdam (EUR), University of Amsterdam (UvA) and Vrije Universiteit Amsterdam (VU).

TI was founded in 1987, and is nowadays among the largest economic research institutes in the Netherlands, with more than 150 research fellows, over 200 PhD candidates and over 770 PhD alumni. Tinbergen Institute is named after Professor Jan Tinbergen, the Dutch Nobel Prize laureate in Economics (1969). TI has offices in both Amsterdam and Rotterdam. TI is located both in Amsterdam and Rotterdam.

The Tinbergen Dissertation Series contains PhD dissertations in the field of Economics, Finance, and Econometrics defended by PhD students that were affiliated to the Tinbergen Institute and were supervised by promotors affiliated to the Tinbergen Institute. For a full list of PhD theses that appeared in the series, I refer to the list of PhD Theses on the following link: https://tinbergen.nl/list-of-phd-theses. The following books recently appeared in the Tinbergen Institute Research Series:

**756 J.H. THIEL**, *Competition, Dynamic Pricing and Advice in Frictional Markets: Theory and Evidence from the Dutch Market for Mortgages*

**757 A. NEGRIU**, *On the Economics of Institutions and Technology: a Computational Approach*

**758 F. GRESNIGT**, *Identifying and Predicting Financial Earth Quakes using Hawkes Processes*

**759 A. EMIRMAHMUTOGLU**, *Misperceptions of Uncertainty and Their Applications to Prevention*

**760 A. RUSU**, *Essays in Public Economics*

**761 M.A. COTOFAN**, *Essays in Applied Microeconomics: NonMonetary Incentives, Skill Formation, and Work Preferences*

**762 B.P.J. ANDRÉE**, *Theory and Application of Dynamic Spatial Time Series Models*

**763, P. PELZL**, *Macro Questions, Micro Data: The Effects of External Shocks on Firms*

**764 D.M. KUNST**, *Essays on Technological Change, Skill Premia and Development*

**765 A.J. HUMMEL**, *Tax Policy in Imperfect Labor Markets*

**766 T. KLEIN**, *Essays in Competition Economics*

**767 M. VIGH**, *Climbing the Socioeconomic Ladder: Essays on Sanitation and Schooling*

**768 YAN XU**, *Eliciting Preferences and Private Information: Tell Me What You Like and What You Think*

**769 S. RELLSTAB**, *Balancing Paid Work and Unpaid Care over the Life-Cycle*

**770 Z. DENG**, *Empirical Studies in Health and Development Economics*

**771 L. KONG**, *Identification Robust Testing in Linear Factor Models*

**772 I. NEAMŢU**, *Unintended Consequences of Post-Crisis Banking Reforms*

**773 B. KLEIN TEESELINK**, *From Mice to Men: Field Studies in Behavioral Economics*

**774 B. TEREICK**, *Making Crowds Wiser: The Role of Incentives, Individual Biases, and Improved Aggregation*

**775 A. CASTELEIN**, *Models for Individual Responses*

**776 D. KOLESNYK**, *Consumer Disclosures on Social Media Platforms: A Global Investigation*

**777 M.A. ROLA-JANICKA**, *Essays on Financial Instability and Political Economy of Regulation*

**778 J.J. KLINGEN**, *Natural Experiments in Environmental and Transport Economics*

**779 E.M. ARTMANN**, *Educational Choices and Family Outcomes*

**780 F.J. OSTERMEIJER**, *Economic Analyses of Cars in the City*

**781 T. ÖZDEN**, *Adaptive Learning and Monetary Policy in DSGE Models*

**782 D. WANG**, *Empirical Studies in Financial Stability and Natural Capital*

**783 L.S. STEPHAN**, *Estimating Diffusion and Adoption Parameters in Networks New Estimation Approaches for the Latent-Diffusion-Observed-Adoption Model*

**784 S.R. MAYER**, *Essays in Financial Economics*

**785 A.R.S. WOERNER**, *Behavioral and Financial Change – Essays in Market Design*

**786 M. WIEGAND**, *Essays in Development Economics*

**787 L.M. TREUREN**, *Essays in Industrial Economics - Labor market imperfections, cartel stability, and public interest cartels*

**788 D.K. BRANDS**, *Economic Policies and Mobility Behaviour*

**789 H.T.T. NGUYEN**, *Words Matter? Gender Disparities in Speeches, Evaluation and Competitive Performance*

**790 C.A.P BURIK**, *The Genetic Lottery. Essays on Genetics, Income, and Inequality*

**791 S.W.J. OLIJSLAGERS**, *The Economics of Climate Change: on the Role of Risk and Preferences*

**792 C.W.A. VAN DER KRAATS**, *On Inequalities in WellBeing and Human Capital Formation*

**793 Y. YUE**, *Essays on Risk and Econometrics*

**794 E.F. JANSSENS**, *Estimation and Identification in Macroeconomic Models with Incomplete Markets*

**795 P.B. KASTELEIN**, *Essays in Household Finance: Pension Funding, Housing and Consumption over the Life Cycle*

**796 J.O. OORSCHOT**, *Extremes in Statistics and Econometrics*

**797 S.D.T. HOEY**, *Economics on Ice: Research on Peer Effects, Rehiring Decisions and Worker Absenteeism*

**798 J. VIDIELLA-MARTIN**, *Levelling the Playing Field: Inequalities in early life conditions and policy responses*

**799 Y. XIAO**, *Fertility, parental investments and intergenerational mobility*

**800 X. YU**, *Decision Making under Different Circumstances: Uncertainty, Urgency, and Health Threat*

**801 G. GIANLUCA**, *Productivity and Strategies of Multiproduct Firms*

**802 H. KWEON**, *Biological perspective of socioeconomic inequality*

**803 D.K. DIMITROV**, *Three Essays on the Optimal Allocation of Risk with Illiquidity, Intergenerational Sharing and Systemic Institutions*

**804 J.B. BLOOMFIELD**, *Essays on Early Childhood Interventions*

**805 S. YU**, *Trading and Clearing in Fast-Paced Markets*

**806 M. GREGORI**, *Advanced measurement and sampling for marketing research*

**807 O. SOONS**, *The past, present, and future of the euro area*

**808 D. G. URZAINQUI**, *The Distribution of Development: Essays on mobility, inequality and social change*

**809 C. PEKER**, *Guess what I think: Essays on the wisdom in meta-predictions*

**810 A. AKDENIZ**, *On the origins of human sociality*

**811 K. BRUTT**, *Strategic interaction and social information*

**812 P. N. KUSUMAWARDHANI**, *Learning Trends and Supply-side Education Policies in Indonesia*

**813 F. CAPOZZA**, *Essays on the Behavioral Economics of Social Inequalities*

**814 D. MUSLIMOVA**, *Complementarities in Human Capital Production: The Role of Gene-Environment Interactions*

**815 J. DE JONG**, *Coordination in Market and Bank Run Experiments*

**816 Y. KIM**, *Micro Studies of Macroprudential Policies Using Loan-level Data*

**817 S TER MEULEN**, *Grade retention, ability tracking, and selection in education*